



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 9806082**

**TO: Manjunath N Rao**  
**Location: 1/10a11 & 10d01**  
**Art Unit: 1652**  
**Monday, July 07, 2003**

**Case Serial Number: 086082**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**CM1-6B02**  
**Phone: 305-9203**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Rao,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

98801

STIC-Biotech/ChemLib

From: Rao, Manjunath N.  
Sent: Wednesday, July 02, 2003 3:15 PM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 10/086,082

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

Date: 7-2-03

Please search the following as soon as possible for application with serial number **10/086,082**

1. SEQ ID NO:1, against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of all results.
2. SEQ ID NO: 2 against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO.

Manjunath N. Rao  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 7/3/03  
Searcher Prep/Review: [Signature]  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 1  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 9300  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

98001 ORFE

**STIC-Biotech/ChemLib**

**From:** Rao, Manjunath N.  
**Sent:** Wednesday, July 02, 2003 3:15 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Sequence search request for 10/086,082

**From:** Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

**Date:** 7-2-03

Please search the following as soon as possible for application with serial number **10/086,082**

1. **SEQ ID NO:1**, against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of all results.
2. **SEQ ID NO: 2** against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO.

Edward Han  
Technical Info. Specialist  
STIC/Biotech  
CM1 6B02 Tel: 305-9205

Point of Contact:  
Linda Basker  
Library Technician  
CM1 1C19 Tel: 308-4500

1-1442 na  
2-328 aa

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 7/13/03  
Searcher Prep/Review: 7/13/03  
Clerical: 6  
Online time: 6

TYPE OF SEARCH: 1  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 03/02  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



# STIC SEARCH RESULTS

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ+Desk





Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpn** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapn** and **.rapn**

***Because they contain data that is confidential, the results of Pending database searches should not be left in the case .***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 15:45:52 ; Search time 25 Seconds  
(without alignments)  
371.181 Million cell updates/sec

Title: us-10-086-082-2

Perfect score: 1748

Sequence: 1 MKWLEES:NAKRGVAGRKPE.....GNVDPKTYVGAMLNKMLIV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177.5	10.2	575	US-08-922-865-2	Sequence 2, Appl
2	177.5	10.2	575	US-08-922-865-2	Sequence 2, Appl
3	91.5	5.2	1848	US-08-922-865-2	Sequence 5, Appl
4	91.5	5.2	1848	PCT-US95-10661A-6	Sequence 6, Appl
5	82.5	4.7	500	US-08-926-842R-64	Sequence 54, Appl
6	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appl
7	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appl
8	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appl
9	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appl
10	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appl
11	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appl
12	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appl
13	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appl
14	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appl
15	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appl
16	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appl
17	80	4.6	628	US-07-952-853-22	Sequence 22, Appl
18	80	4.6	628	US-07-952-853-22	Sequence 22, Appl
19	79.5	4.5	1544	US-08-914-848-22	Sequence 46, Appl
20	79	4.5	374	US-09-306-446C-2	Sequence 2, Appl
21	79	4.5	1342	US-08-484-438-9	Sequence 9, Appl
22	79	4.5	1711	US-08-369-822C-10	Sequence 10, Appl
23	79	4.5	1711	US-08-369-822C-10	Sequence 10, Appl
24	79	4.5	1711	US-08-369-822C-10	Sequence 10, Appl
25	78.5	4.5	375	US-08-434-831B-10	Sequence 7, Appl
26	78	4.5	383	US-08-391-916A-4	Sequence 4, Appl
27	78	4.5	626	5268290-2	Patent No. 5268290

28	77.5	4.4	1170	US-08-789-078-2	Sequence 2, Appl
29	77.5	4.4	1170	US-08-789-078-2	Sequence 2, Appl
30	77.5	4.4	1170	US-08-789-078-2	Sequence 2, Appl
31	77	4.4	565	PCT-US95-04886-2	Sequence 11, Appl
32	77	4.4	565	US-09-142-623-11	Sequence 7, Appl
33	76.5	4.4	225	US-08-107-884H-10	Sequence 10, Appl
34	76.5	4.4	225	US-08-107-884H-11	Sequence 21, Appl
35	76	4.3	456	US-09-358-364-21	Sequence 4394, Ap
36	76	4.3	629	US-09-134-001C-4394	Sequence 9, Appl
37	76	4.3	1065	US-08-630-172-9	Sequence 9, Appl
38	76	4.3	1065	US-09-375-419-9	Sequence 5106, Ap
39	76	4.3	1177	US-09-134-001C-5106	Sequence 15, Appl
40	76	4.3	1222	US-08-682-517-15	Sequence 9, Appl
41	76	4.3	1252	US-08-682-517-9	Sequence 4, Appl
42	76	4.3	1346	US-09-320-878-4	Sequence 37, Appl
43	76	4.3	1346	US-09-105-537-37	Sequence 6, Appl
44	76	4.3	11877	US-09-105-537-6	Sequence 12, Appl
45	75.5	4.3	354	US-09-500-569-12	

## ALIGNMENTS

RESULT 1  
US-08-922-865-2  
Sequence 2, Application: US/08922865  
Patent No. 6050616  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: MICROORGANISM, LACTAMASE ENZYME OBTAINED  
THEREFROM, AND THEIR USE  
NUMBER OF SEQUENCES: 2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,865  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-922-865-2

Qy	39	PSYTP-----VLTEPGDRIVUTRAFEAGI---NSEQDIPS-GILKHPFLN 82	Query Match	10.2%	Score 177.5	DB 3	Length 575
Qy	44	PYCNQVNRHWFETPMMAVVEPCAEFLTYDWTGCAIKNDOSAEIVRDVLSVWFL 102	Best Local Similarity	24.0%	Prod. No. 3e-11		
Qy	83	PONGPIMVNAKGVIVLAV-----VIESMPEGV-----C 112	Matches	59	Conservative	33	Mismatches 85
Qy	103	--SGPVGVKGAQPGCLLVLLDISAKDCLSWGFNGFFSKONGGFLDEHFFLAKRS:WD 160					
Qy	113	PGYICAMTPH-----FGGLTGTDLTAMNOLPEKVKMIKLDSEKVVWKRHFLYKPHI 167					
Qy	161	PHOMTEKSRHFGVNFAGLHFGGLICLPD-----KML-----ASWNERET ----- 202					
Qy	165	GTLSVSPSE-----IDTNSLTPDNHSGNDMDVPGIGNG 194					
Qy	203	GLIATDPRIPGLANPPNATTAHMQOMGEAKKAAAGARTVPPPHGSGND:KOLSRG 262					
Qy	200	SITYPLVRAPGSKLFGIDAHACAGDGEIC-GTAVEFASITTIKVDLIK 246					
Qy	263	SRVFPVYVDGAGLSVGLHFSQGDGELTFWGPPIEMPGWIMKVSUIK 310					

RESULT 2  
US-09-510-949-2  
; Sequence 2, Application US/09510949  
; Patent No. 6423522  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene III, Joseph W.  
; TITLE OF INVENTION: MICROORGANISM, LACTAMASE ENZYME OBTAINED  
; NUMBER OF SEQUENCES: 2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/510,949  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 06/922,655  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 575 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-510-949-2

Query Match 10.2%; Score 177.5; DB 4; Length 575;  
Best Local Similarity 24.0%; Pred. No. Re-11;  
Matches 69; Conservative 33; Mismatches 85; Indels 101; Gaps 12;  
QY 39 PYSTP-----VLTIEGDRILVCTRDFAEQA-----NSKODIPS-QLLKMPFLN 82  
DB 44 PYDNFOVHRWIPDIPMAVWVEPGAEEFKLETITWTGGA:KNDSABEDVDLSVTFE- 122  
QY 83 PONGPIKVNAGKGDVLAV-----YIFSLMIPROV-----E 112  
DB 103 --SGPVGKAGQGLLVVLDLDIGARDLSLWGFNGFSGKQNGGFLDEHFPLAGK:WD 160  
QY 113 PYGICAMIPH-----FGLGIGLIDLAMLNDPLPEKVRMIKLDSEKVVYKSHLTPYKPHI 167  
DB 161 FHOMFTKSRHPGVNPAFLIHPGLIGCLPDP-----KML-----ASNNREF----- 202  
QY 168 GLTSLVSE-----IDSLNLTPOHNGNDVDPDIPG 199  
DB 203 GLIATDPDRIPGLANPPNATTIAHGMOMGEAKDAASGARTVPPREHGGNCIDKLSRG 262  
QY 200 SITPLVRAFGGLFGLGDAHACQGDGEIC-GTAVEFASITTIKYDLIK 246  
DB 263 SRVFFPVYVDGAGLSVGDJHFSGDGEITFMGPIEMPGWVHKVSLIK 310

RESULT 3  
US-06-296-791-6  
; Sequence 6, Application US/08296791  
; Patent No. 6245337  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene III, Joseph W.  
; APPLICANT: Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/REF/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277295  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1848 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
US-06-296-791-6

Query Match 5.2%; Score 91.5; DB 4; Length 1848;  
Best Local Similarity 19.1%; Pred. No. 1.4;  
Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 20;  
QY 35 YTIQPYSTPVLTEPGD-----RIIVCTRDFAE-GAINSE-ODIPSOLLKMFLLNPONGPI 88  
DB 17 YALPYTEARLVROVDYQIFRDPAENKGFESVGA:NVFDRKKNSL-----GSA 67  
QY 89 MVNAGKGDVLAVYTESMLPGEVDPY-----GICAMIFHFGGLGTITLAKENP 138  
DB 68 LEPGTPMIDFVSFVDEKRIATLVNPQYVGVGVKHYKNSVSEL--HFGKLNCGN-----MNG 120  
QY 139 LPEKVM:KLDSKVVYKSKRHTLP-----YKPHLCTLSVSEFID 177  
DB 121 NAKSHRQVSSSENFYVVEKNNFETNVTFTKEQDAKRRREDYTMPRLDKFVT--EVA 178  
QY 178 SINSLTPONHNGNDVDPDIPGSGITP-LVRAFGGLFGLGDAHACQGDGEICGTAVEFAS 236  
DB 179 P-EASTANNKGEYNSD-----KYPAFVRLSGTGFIYKGS-----RYQL 220  
QY 237 ITTIK-----VDIKKNQLSNURME-----NAENIMSIGSARPLE--DA 273  
DB 222 ILTEKDKGGLIRNWDGGDLVLGNATYGTIAGTPYKVNHNHNSL:KFGNKEEESDH 287  
QY 274 TRAYRPD-----LIY-----WLV-----EDF-----GFEQWYMLLSQ 302  
DB 281 KGLISQDPLTNVAVLSGSSPLFVYDREKGNKFLGSDYDFWAGYKNSKSWENNY----- 335  
QY 303 COKVRLGNWDFKTYTVGAMLNKN 325  
DB 336 --KHEFAEKLYQOVSAGSLIGSN 356

RESULT 4  
PCI-US95-10661A-6  
; Sequence 6, Application PC/TUS9510661A  
; GENERAL INFORMATION:  
; APPLICANT: Washington University, et al.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/10661A  
 ; FILING DATE: 16-AUG-1995  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/296,791  
 ; FILING DATE: 25-AUG-1994  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Trecartin, Richard F.  
 ; REGISTRATION NUMBER: 31,801  
 ; REFERENCE/DOCKET NUMBER: FP-59941/REF  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 791-1989  
 ; TELEFAX: (415) 398-3249  
 ; TELEX: 910 277295  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1848 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; PCT-US95-10661A-6

Query Match 5.2%; Score 91.5; DB 5; Length 1848;  
 Best Local Similarity 19.1%; Pred. No. 1.4;  
 Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 20;  
 QY 35 YTGPGVSTPVLTTEPGD---RIIVDIRDAFE-GAINSE-QDIPSOQLKMPFLNPNQNGE 48  
 DB 17 YALTPYTEALVRODDVDYQIFRDAENKGFSGATNVEVRKKKNSL-----GSA 67  
 QY 89 MYNGAEKGDVLAIVTESMPLRGVDPY-----GICAMIRFEGGLTGLTGLTAMNDP 138  
 DB 68 LPNGIPMIDFVSVDVQKRIATLNPQYVGVKHSVNGVSEL--HFGNLNGN-----MNG 120  
 QY 139 LPEKVRMIKLDSEKVMYKSRHTLP-----YKPHIGTSLVSPEID 177  
 DB 121 NAKSHRDVSEENRYITVENKNPTEIVTSFKKEOCACKREDIYKPKDFVT--EVA 178  
 QY 178 STNLTDPNCGNNDVPDIPGSIITYP-LVRAPGGRLFGDAHACQDSECGTAVEFAS 236  
 DB 179 PIEASTANNKGEYNND-----KYPAFVRIGSOTQFYKKS-----RYQL 220  
 QY 237 ITLIK---VDLIKMWLSKPRME-----NAFNIMISGASRQLE--DA 273  
 DB 221 LLEKKGKGLLRMDVGGNLELVGNVYTYGIAGTYKVNHNKGLIGFGRSKERSHP 280  
 QY 274 TRAYRD-----LIY---WLV---EDF-----GFEQKDAYMLLSQ 302  
 DB 281 KGLSQDPLTNVAVLGDGSPLFVYDREKCKWIFLGSYDFWAGYKKNKSNCEWNIY----- 315  
 QY 403 CGKVRLGMMVDPKYTVGAMLNK 325  
 DB 336 --KHEFAEKIVQOYSAGSLGNS 356

RESULT 5  
 US-08-926-842B-64  
 ; Sequence 64, Application US/08926842B  
 ; Patent No. 6030807  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sa-No. 6030807, Leira, Isabel  
 ; APPLICANT: de Lencastre, Herminia  
 ; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE  
 ; TITLE OF INVENTION: EXPRESSION  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Klauder & Jackson  
 ; STREET: 411 Hackensack Avenue  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/926.842B  
 ; FILING DATE: 10-SEP-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Ess., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 603-1-089 N  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201 487-5860  
 ; TELEFAX: 201 343-1584  
 ; TELEX: 133521  
 ; INFORMATION FOR SEQ ID NO: 64:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 500 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Protein  
 ; ORIGINAL SOURCE: Bacillus subtilis  
 ; ORGANISM: Bacillus subtilis  
 ; FEATURE:  
 ; OTHER INFORMATION: /product- "abta"  
 ; US-08-926-842B-64  
 Query Match 4.7%; Score 82.5; DB 3; Length 500;  
 Best Local Similarity 25.4%; Pred. No. 1.7;  
 Matches 51; Conservative 32; Mismatches 75; Indels 43; Gaps 13;  
 QY 141 EKVRMIKLDSE-----KVYMSKRHTLPYKPHIGTSLVSPEIDSLTPDHHGGMND 192  
 DB 2 KKARMI-VCKEKIGKGVKRIYGS-----FIEHMGK-AVTEGIVEPDRPEADEGGERKD 53  
 QY 193 VPDIIGSGITTYPLVAPGGRLFGDAHACQDGEICG--TAVEFAS:TTIKVCLIKMWCL 250  
 DB 54 VQSL-TKEQVPLIRYPGGN-FLSGYNWEDGVGVENRPRDLAWQTE-NEVG:NEFL 11;  
 QY 251 SWPMENALNTMSIG-SASPLEDATRIAYEDLI-----YV--LVEDGFQGWDAYN 296  
 DB 112 SWAKKVNTEVMAVNLGTSID-----AARNLVEYCNPKGSYWDJKRSHGYED--PYG 164  
 QY 299 LLSQCKKVRPLGNVTPKRYTVG 319  
 DB 165 IKTWC-----LGNEMDGFQWIG 181  
 RESULT 6  
 US-08-779-113-2  
 ; Sequence 2, Application US/08779113  
 ; Patent No. 5948891  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Staunton, Donald E.  
 ; APPLICANT: Harris, Edith S.  
 ; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin  
 ; TITLE OF INVENTION: Binding  
 ; NUMBER OF SEQUENCES: 65  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 233 South Wacker Drive, 6300 Sears Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/03/779,113

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Greta E. No. 594867land

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/33773

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 857 amino acids

TYPE: amino acid

TOPOLOGY: not relevant

MOLECULE TYPE: protein

US-03-779-113-2

Query Match  
Best Local Similarity 20.3%; Pred. No. 4.3;  
Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 SESIMAKR-----GVGAGRKPVTHLTEMQKEHYTHIGPSTPVLTIEFGGR-----52  
DB 204 DEPCILKRLPNNHIGISLIPREVGEHVS:KKNGHNVANSFVSINVQSEIGDARRAKVY 263  
QY 53 -----LIIVTRDAFEAGINSEODIPSOILKMPFLNPONGPIMVNGAEKGDV 50  
DB 264 GRLSGRTFMSDFIVDTIDAGYGGIS-----LAVGPGSKVDI 302  
QY 95 LAVYIESMLPSGVDPGICAMIPHEGGLTGIDL--TAMLNOLPEKVKIKLDSKVVYS 156  
DB 303 QTEDELD-----GTC-KVSYPTVPGVYIVSTKXZADEHVPSPPTVKISGEG---348  
QY 157 KRHTLPYKPHIGTISVSPRIEISINSITPONHGNMDVPDIPGSGTYPPLVRAFGGR 213  
DB 349 -----RVKES:TRTSRAPSVATVGSIC-----DLNLKIPINSSDMS-AHVTSPSGRV 395

#### RESULT 7

US-08-583-562B-2

Sequence 2, Application US/08583562B

Patent No. 5922570

GENERAL INFORMATION:

APPLICANT: Staunton, Donald

APPLICANT: Harris, Edith

TITLE OF INVENTION: Cytoplasmic Modulators of Integrin

TITLE OF INVENTION: Binding

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/583,562B

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/33038

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 958 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-583-562B-2

Query Match 4.7%; Score 82.5; DB 2; Length 958;

Best Local Similarity 20.3%; Pred. No. 4.3;

Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 SESIMAKR-----GVGAGRKPVTHLTEMQKEHYTHIGPSTPVLTIEFGGR-----52  
DB 204 DEPCILKRLPNNHIGISLIPREVGEHVS:KKNGHNVANSFVSINVQSEIGDARRAKVY 263  
QY 53 -----LIIVTRDAFEAGINSEODIPSOILKMPFLNPONGPIMVNGAEKGDV 50  
DB 264 GRLSGRTFMSDFIVDTIDAGYGGIS-----LAVGPGSKVDI 302  
QY 95 LAVYIESMLPSGVDPGICAMIPHEGGLTGIDL--TAMLNOLPEKVKIKLDSKVVYS 156  
DB 303 QTEDELD-----GTC-KVSYPTVPGVYIVSTKXZADEHVPSPPTVKISGEG---348  
QY 157 KRHTLPYKPHIGTISVSPRIEISINSITPONHGNMDVPDIPGSGTYPPLVRAFGGR 213  
DB 349 -----RVKES:TRTSRAPSVATVGSIC-----DLNLKIPINSSDMS-AHVTSPSGRV 395

#### RESULT 8

US-09-556-877-180

Sequence 180, Application US/09556877

Patent No. 6432916

GENERAL INFORMATION:

APPLICANT: Probst, Peter

APPLICANT: Bralio, Ajay

APPLICANT: Skeiky, Yasir

APPLICANT: Fling, Steve

APPLICANT: Maisonneuve, Jeff

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

FILE REFERENCE: 210121.459C5

CURRENT APPLICATION NUMBER: US/09/556.877

CURRENT FILING DATE: 2000-04-19

NUMBER OF SEQ ID NOS: 305

SOFTWARE: FASTSEQ for Windows Version 3.0/4.0

SEQ ID NO 180

LENGTH: 1752

TYPE: PRT

ORGANISM: Chlamydia

US-09-556-877-180

Query Match 4.7%; Score 82.5; DB 4; Length 1752;

Best Local Similarity 23.1%; Pred. No. 14;

Matches 49; Conservative 26; Mismatches 60; Indels 77; Gaps 10;

QY 40 YSTPVLTIETGRIIVTTRDAFE--GAINSEODIPSOILKMPFLNPONGPIMVNGAE---94  
DB 271 YAGELTIADSQEVFFSINKATOGGAIPAEDV-----SFENITSLKVATNGAEKGS 423  
QY 95 -----KGDVL-----AVYIESMLPRGVDPPYGCICAMIPHEGGLTGID 130  
DB 324 CALYAKHDLISQSSKOSLNSNYSKGGALVE-----SGINPQD 364  
QY 131 LTAMNOLPEKVKIKLDSKVVYSKRNITLPYKPHIGTISVSPIDISINSLTPONHGN 190  
DB 365 L-----EELP-IRYNKAGTETKTKIT---PSLKAQAGAGNADAWASSSSVQSSGA 419  
QY 151 MCVPD:GPGS!-----TYP:VRAFGGRKLF 213  
DB 412 TVVSDSSDSSSGSDSTSETVP-VTAKGGGLY 442

#### RESULT 9

```

US-09-620-412C-180
: Sequence 180, Application US/09/26412C
: Patent No. 6448234
: GENERAL INFORMATION:
: APPLICANT: Steven P. Fling
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: FILE REFERENCE: 210121.469C7
: CURRENT APPLICATION NUMBER: US/09/620.412C
: CURRENT FILING DATE: 2000-07-20
: NUMBER OF SEQ ID NOS: 363
: SOFTWARE: FastSeq for Windows Version 3.0/4.0
: SEQ ID NO 180
: LENGTH: 1752
: TYPE: PRT
: ORGANISM: Chlamydia
US-09-620-412C-180

Query Match      4.7%  Score 82.5; DB 4; Length 1752;
Best Local Similarity 23.1%; Pred. No. 14;
Matches 46; Conservative 26; Mismatches 60; Indels 77; Gaps 10;

QY 40 YSTPVLTIERGDIIVDTDAPE--GAINSEDDIPSOILKMFPLNPQNGPIMVNGAE--- 94
DB 271 YAKGDLIADSOEVSINKATKGGAIKAEKV-----SFENITSLKVGINGAEKGG 323
QY 95 -----KGDVL-----AVYIESMLPRGVDPTGTCAMIFHFGSLGTGD 130
DB 324 GAIYAKGDLISQSSKOSLNSYNSKOGGALAYE-----GGINPQD 364
QY 131 ITAMLNPLPEKVMKLDSEKVKYKSRHTLPYKPHLGTLSVSFEIUSNSLTPNGHGN 140
DB 365 L-----EIR-IRYNKAGTFEKKITL---PSLKACAGNADAWASSPQNGSA 111
QY 191 MVDVDPGCSI-----TYPVLRAPGGKLF 214
DB 412 TVVSQSDSSSGSDUSTSETVF-VTAKGGGLY 442

RESULT 10
US-09-177-165A-30
: Sequence 30, Application US/29177155A
: Patent No. 6426205
: GENERAL INFORMATION:
: APPLICANT: Tyers, Mike
: APPLICANT: Williams, Andrew
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING ERICUTIN
: FILE REFERENCE: 11757.10USU1
: CURRENT APPLICATION NUMBER: US/09/177.165A
: CURRENT FILING DATE: 1998-10-22
: PRIOR APPLICATION NUMBER: 60/092.443
: PRIOR FILING DATE: 1998-07-10
: PRIOR APPLICATION NUMBER: 60/045.254
: PRIOR FILING DATE: 1997-10-24
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 30
: LENGTH: 640
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-30

Query Match      4.7%  Score 82; DB 4; Length 640;
Best Local Similarity 21.1%; Pred. No. 3;
Matches 39; Conservative 23; Mismatches 65; Indels 58; Gaps 7;

QY 82 NPQNGPIMVNGAEKGEVLAVYIESMLPRGVDPTGTCAMIFHFGSLGTGTJAMCNPLPE 141
DB 450 SPQDDPTMDCADESDFPSNEQEVJDE-----NTPYPT 523
QY 142 KVMKLDSEKVKYKSRHTLPYKPHLGTLSVSFIDSNSLTPGN-----HGGNNDVP 194

```

```

DB 524 HLLSCGLDNTKLVNKTGKIRIQFG-----EVEGVWDIAADNFRILSGSHSSIKKW 577
QY 195 DISPCSTITYELVRAPGGRNL-----FQGD-----AHACOGEGHIGTAVEFASITT 239
DB 578 DLOGSKCMHTF---NORRLORETQHTQTSGLGDKVAPIACVCIQDSE-CFSGDGRGVAK 633
QY 240 IKVDL 244
DB 534 YKFDL 638

RESULT 11
US-08-476-062A-42
: Sequence 42, Application US/08476062A
: Patent No. 587275
: GENERAL INFORMATION:
: APPLICANT: Arndout, M. Amir
: TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
: RESPONSES WITH BETA2 INTEGRINS
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476,062A
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 05/216,081
: FILING DATE: 21-MAR-1994
: APPLICATION NUMBER: 07/637,830
: FILING DATE: 04-JAN-1991
: APPLICATION NUMBER: 07/535,842
: FILING DATE: 18-JUN-1990
: APPLICATION NUMBER: 07/212,573
: FILING DATE: 28-JUN-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Freeman, John W.
: REGISTRATION NUMBER: 29,066
: REFERENCE/DOCKET NUMBER: 00786/066003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 42:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1170 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
US-08-476-062A-42

Query Match      4.7%  Score 81.5; DB 2; Length 1170;
Best Local Similarity 22.7%; Pred. No. 9.4;
Matches 70; Conservative 35; Mismatches 109; Indels 95; Gaps 18;

QY 14 VGARKKPVTHLTIEEMOKERH-YTIGPYSTPVLTIPEGDRIVDTDAFEGAINSEQDIP 72
DB 284 IGIGK---HFQIKESQETLHKFKASKPASE-----FKILDIEF----- 315
QY 73 SOLLKMPFLNPNQNGPIMVNGAEKGEVLAVYIE-----SNLPRGVDPYKCAKIPHFG 125
DB 319 --KLKDLFIERQKKIYVIEGTSKQCLTSFNMEI.SSSGISADLSRGNVAVGAKKAGG 376

```

QY 126 LTGTLTAMC-----NDPLPEKVRMIKLDSEKVVW--SKRST-----LFXPHLQTE 172  
DB 377 F--LDLKAELQDDTFIGNETPTPEVRAGYL-GYTVTWLPSRQKTSLLASGAPFYOHMGV 433  
QY 171 SVSPE-----IUSNSLTPONH-GGNKVPVJIG-POSITYPLVRAP-----GSR 212  
DB 434 LLFQPOGGCHWSQVOTHTGTSFGGELCGVDVQDGTETELLIGAPFYGQRGR 493  
QY 213 LFI-----GDAHACQSD-GEICUTAVEFASITIKVDLIKNNQLSWPRMENAEINHS 263  
DB 494 VFIYORRQIGFEFVSELQDGPVFLG---RFGESAITALTJ1-----NKGGLVD 538  
QY 264 IGSARPLED 272  
DB 539 VAVGAPLEE 547  
RESULT 12  
PCT-US96-01114-42  
Sequence 42, Application PC/TUS9601314  
GENERAL INFORMATION:  
APPLICANT: M. Amin Arnaut  
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN  
TITLE OF INVENTION: ANTAGONISTS  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01314  
FILING DATE: 30-JAN-96  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/380,167  
FILING DATE: 30-JAN-95  
ATTORNEY/AGENT INFORMATION:  
NAME: John W. Freeman  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: G0786/267001  
TELEPHONE: (617) 542-5076  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1170  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US96-01114-42

Query Match 4.7% Score 82.5; DB 5; Length 1170;  
Best Local Similarity 22.7%; Pred. NO. 9.4;  
Matches 70; Conservative 35; Mismatches 105; Indels 95; Gaps 18;

QY 14 VAGRKPVTHHTEEMQKEFH-YTICPVSTPVLITIEPCDRIIVTRAFEGAINSEQDIP 72  
DB 284 IGIGK-----HFQTKESQETJHFKASKPASE-----FVKILDTF----- 316  
QY 73 SOLIKMPLPONGPIMVNGAEKGEVLAVYIP-----SMIPKSYDVPYGTAMIIHFGS 127  
DB 319 --ALKDLFTERQKKIIVLECTSKQDILTSFNMESSSGISACLSRCHAVGVAGAKWAG 375  
QY 126 LTGTDLTAMC-----NDPLPEKVRMIKLDSEKVVW--SKRST-----LFXPHLQTE 170

DB 377 F--LDLKAELQDDTFIGNETPTPEVRAGYL-GYTVTWLPSRQKTSLLASGAPFYOHMGV 433  
QY 171 SVSPE-----IUSNSLTPONH-GGNKVPVJIG-POSITYPLVRAP-----GSR 212  
DB 434 LLFQPOGGCHWSQVOTHTGTSFGGELCGVDVQDGTETELLIGAPFYGQRGR 493  
QY 213 LFI-----GDAHACQSD-GEICUTAVEFASITIKVDLIKNNQLSWPRMENAEINHS 263  
DB 494 VFIYORRQIGFEFVSELQDGPVFLG---RFGESAITALTJ1-----NKGGLVD 538  
QY 264 IGSARPLED 272  
DB 539 VAVGAPLEE 547  
RESULT 13  
US-07-978-895-4  
Sequence 4, Application US/0797895  
Patent No. 5480968  
GENERAL INFORMATION:  
APPLICANT: Kraus, Mathias H.  
APPLICANT: Aaronsen, Stuart A.  
TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE  
TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND  
TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sullig 400  
STREET: 133 Carnegie Way, N.W.  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.A.  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/978,895  
FILING DATE: 19921110  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/444,406  
FILING DATE: 01-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Perryman, David S.  
REGISTRATION NUMBER: 33,435  
REFERENCE/DOCKET NUMBER: 1414-029  
TELEPHONE: (404) 688-0770  
TELEFAX: (404) 688-9560  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1342 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-978-895-4

Query Match 4.6% Score 81; DB 1; Length 1342;  
Best Local Similarity 24.6%; Pred. No. 13;  
Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY 5 IMAKRGVAGRKPVTHHTEEMQKEFH-YTICPVSTPVLITIEPCDRIIVTRAFEGAINSE 67  
DB 978 KRESSPGTANGFETHSLTKKK:EEVE--DEPEDLDLDUEAEEDNLT1-----TDS 1029  
QY 98 LODIPSOLEKMP-----FLAFONGPIMVNGAEKGEVLAVYIESKL-----PKGYD--- 112  
DB 1030 ALSLVGTLNPRGQSLSLSPSSGYMPMNGNLGSCG---PSAVGSSESGCPRPVSLHP 1096  
QY 113 -PYGICAMIPFGSLTGILTLAKNDLPKVRMIKLDSEKVVW--SKRST-----EKKVSKKHT1- 161

```

Db      1387 MPROCLASSESEGVITGSEA-----ELQEKVSMCRSRSSRSPRPGDSAYHSORHSL 1140
QY      162 -PYKPHIGTISVSP---EIDINS-LTPDNH 187
DB      1141 TPVTP-----LSPGLEDVNGYVMPDTH 1165

RESULT 14
US-08-473-119-4
: Sequence 4, Application US/08473119
: Patent No. 5820859
: GENERAL INFORMATION:
: APPLICANT: Kraus, Matthias H.
: APPLICANT: Aaronson, Stuart A.
: TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
: TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND
: TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THERETO
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Suite 400
: STREET: 133 Carnegie Way, N.W.
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: U.S.A.
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.C, Version #1.25
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/473,119
: FILING DATE: 07-JUN-1995
: CLASSIFICATION:
: APPLICATION NUMBER: 07/978,895
: FILING DATE: 10-NOV-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Perryman, David G.
: REGISTRATION NUMBER: 33,438
: REFERENCE/DOCKET NUMBER: 1414-028
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 688-0770
: TELEFAX: (404) 688-9880
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1342 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-473-119-4

Query Match: 4.6%; Score 81; DB 2; Length 1342;
Best Local Similarity 24.6%; Pred. No. 13;
Matches 52; Conservative 26; Mismatches 75; Indels 54; Gaps 12;

QY      8 IMAKRGVAGKPKVTHLITEEMCKEFHYTIGPYSTPVLTIETPGDRIIVDTRCAFESAINS 67
DB      978 IKRESSGCIAPGPEPHGLTNKKLEVE--LPDLDDLDEAEEDNLAT-----TLGS 1029
QY      68 EODIPSQLLKMP-----FLNPQNGPIVNGAEKGVLAIVIESML-----PRGVD--- 112
DB      1030 ALSLPVGTLLNPRGSQLSLSSGYMPMQNLGESCQ---ESAVSGSSERCPRPVSLHP 1066
QY      113 -PYGICAMPHFGGLTGTDLTAMLDNLPKVKRMKILDS-----EKVYNSKRHTL- 161
DB      1087 MPRGLASESSEGHVTGSEA-----ELQEKVSMCRSRSSRSPRPGDSAYHSORHSL 1140
QY      162 -PYKPHIGTISVSP---EIDINS-LTPDNH 187
DB      1141 TPVTP-----LSPGLEDVNGYVMPDTH 1165

```

Search completed: July 3, 2003, 15:50:45  
Job time : 38 secs

```

DB      1141 TPVTP-----LSPGLEDVNGYVMPDTH 1165

RESULT 15
US-08-475-352-4
: Sequence 4, Application US/08475352
: Patent No. 5916755
: GENERAL INFORMATION:
: APPLICANT: Kraus, Matthias H.
: APPLICANT: Aaronson, Stuart A.
: TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
: TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND
: TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THERETO
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Suite 400
: STREET: 133 Carnegie Way, N.W.
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: U.S.A.
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.G, Version #1.25
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,352
: FILING DATE:
: CLASSIFICATION:
: APPLICATION DATA:
: APPLICATION NUMBER: 07/978,895
: FILING DATE:
: APPLICATION NUMBER: US 07/444,406
: FILING DATE: 01-DEC-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Perryman, David G.
: REGISTRATION NUMBER: 33,438
: REFERENCE/DOCKET NUMBER: 1414-028
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 688-0770
: TELEFAX: (404) 688-9880
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1342 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-475-352-4

Query Match: 4.6%; Score 81; DB 2; Length 1342;
Best Local Similarity 24.6%; Pred. No. 13;
Matches 52; Conservative 26; Mismatches 75; Indels 54; Gaps 12;

QY      8 IMAKRGVAGKPKVTHLITEEMCKEFHYTIGPYSTPVLTIETPGDRIIVDTRCAFESAINS 67
DB      978 IKRESSGCIAPGPEPHGLTNKKLEVE--LPDLDDLDEAEEDNLAT-----TLGS 1029
QY      68 EODIPSQLLKMP-----FLNPQNGPIVNGAEKGVLAIVIESML-----PRGVD--- 112
DB      1030 ALSLPVGTLLNPRGSQLSLSSGYMPMQNLGESCQ---ESAVSGSSERCPRPVSLHP 1086
QY      113 -PYGICAMPHFGGLTGTDLTAMLDNLPKVKRMKILDS-----EKVYNSKRHTL- 161
DB      1087 MPRGLASESSEGHVTGSEA-----ELQEKVSMCRSRSSRSPRPGDSAYHSORHSL 1140
QY      162 -PYKPHIGTISVSP---EIDINS-LTPDNH 187
DB      1141 TPVTP-----LSPGLEDVNGYVMPDTH 1165

```



\_\_\_\_\_

SeqCore version 5.1.6  
Copyright (c) 1993 - 2003 Computer Ltd.

OM nucleic - nucleic search, using sw mode:

Run on: July 6, 2003, 04:04:05 : Search time 1972 seconds

(without alignments)  
11842.744 Million cell: updates/sec

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtggcc.....aatgaattcatttgatccc 1442

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 42302132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:

1: em\_estb1:\*  
2: em\_estm1:\*  
3: em\_estm2:\*  
4: em\_estm3:\*  
5: em\_estm4:\*  
6: em\_estm5:\*  
7: em\_estm6:\*  
8: em\_estm7:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estm1:\*  
16: em\_estm2:\*  
17: em\_estm3:\*  
18: em\_estm4:\*  
19: em\_estm5:\*  
20: em\_estm6:\*  
21: em\_estm7:\*  
22: em\_estm8:\*  
23: em\_estm9:\*  
24: em\_estm10:\*  
25: em\_estm11:\*  
26: em\_estm12:\*  
27: em\_estm13:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.6	2.7	280	10 AW380638	AW380638 RC2-HT027
2	37.8	2.6	551	13 BM615374	BM615374 170006871
3	36.2	2.5	533	10 AV551507	AV551507 AV551507
4	36.2	2.5	678	13 BM582997	BM582997 170006872
5	36.2	2.5	1101	17 CNS00370	ALC64465 Drosophila
6	36	2.5	495	10 AV629284	AV629284 AV629284

C	7	35.6	2.5	482	17	BM767573
C	8	35.6	2.5	735	17	AG117547
C	9	35.2	2.4	612	10	BE442656
C	10	35.2	2.4	772	17	BM704682
C	11	35	2.4	268	9	AA020733
C	12	35	2.4	421	14	H75593
C	13	35	2.4	441	9	AA019639
C	14	35	2.4	452	9	AA061557
C	15	35	2.4	505	12	BC834211
C	16	35	2.4	531	17	B57961
C	17	34.8	2.4	528	9	AG109318
C	18	34.8	2.4	396	17	AG013952
C	19	34.8	2.4	422	17	AG055596
C	20	34.8	2.4	475	9	AT43670
C	21	34.8	2.4	477	9	AT436857
C	22	34.8	2.4	482	14	BC247372
C	23	34.8	2.4	502	9	AT289776
C	24	34.8	2.4	524	12	RG052847
C	25	34.6	2.4	511	17	AG048269
C	26	34.6	2.4	562	17	AG048968
C	27	34.4	2.4	410	12	BF484542
C	28	34.4	2.4	474	10	BF498530
C	29	34.4	2.4	483	9	AG033827
C	30	34.4	2.4	491	14	H83687
C	31	34.4	2.4	623	17	B59197
C	32	34.4	2.4	641	13	BM392441
C	33	34.4	2.4	641	13	BM393654
C	34	34.4	2.4	100	17	CNS00020
C	35	34.2	2.4	466	13	BM136789
C	36	34.2	2.4	531	13	BM246954
C	37	34.2	2.4	656	13	BM234037
C	38	34.2	2.4	681	10	BF489900
C	39	34.2	2.4	523	13	BM018122
C	40	34	2.4	347	14	RG3461
C	41	34	2.4	515	12	BF252201
C	42	34	2.4	540	12	BF252018
C	43	34	2.4	565	17	AG043850
C	44	34	2.4	569	12	BF251067
C	45	34	2.4	750	12	RG574169

#### ALIGNMENTS

RESULT 1  
AW380638  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
Human.  
Homo sapiens  
Pekaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RC2-HT0275-01-199-011-11 HT0275 Homo sapiens cDNA, mRNA sequence.  
AW380638  
AW380638.1 GI:5885297  
EST.  
EST 04-FEB-2000

REFERENCE  
1 (bases 1 to 280)  
The FAPESP/LICR Human Cancer Genome Project  
Unpublished (1999)  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: -55-11-2704922  
Fax: -55-11-2707000  
Email: asimpson@ludwig.org.br

PROJECT  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC2&t2=RC2-HT0275-01199-011-flt&t3=1999-11-01&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 37



```

RESULT 4
LOCUS      BM582997
DEFINITION 1700068727733 A.Gam.ad.cdna.blood1 Anopheles gambiae cDNA clone
ACCESSION 19600449689205 5', mRNA sequence.
VERSION    BM582997
KEYWORDS   EST, BM582997.1 GI:18871464
SOURCE     African malaria mosquito.
ORGANISM   Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
            Anopheles.
REFERENCE  1 (bases 1 to 678)
AUTHORS   Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Chariab,
            R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
TITLE     CelerA Anopheles gambiae EST project
JOURNAL   Unpublished (2002)
COMMENT   Contact: Holt R.A.
            CelerA Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltR@cclera.com
            Plate: NU01004Y0 row: G column: 15
            Seq primer: M13 Reverse.
            Location/Qualifiers
                1..678
                /organism="Anopheles gambiae"
                /strain="RSP-ST (Reduced suscep. to Permethrin - std.
                chromosome)"
                /db_xref="taxon:7165"
                /clone="19600449689205"
                /clone_lib="A.Gam.ad.cdna.blood1"
                /dev_stage="Adult"
                /lab_host="Drl0b"
                /note="Vector: pSport1; Site1: SalI; Site2: NotI; Whole
                adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                hours after human blood feeding. cDNA insert: 6500 bp
                cloned directionally into pSport 1. Not 1 site is 3'.
                Clones available through the Malaria Research and
                Reference Reagent Resource Center (www.malaria.rrc.ri.org)"
            BASE COUNT 177 a 178 c 228 g 95 t
            ORIGIN
                10 TCGATGCGCGGTGATGCTGTCGAGGAGGATATGAGATGATGAGCGGCGGACAGC 69
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 261 TCCAGCGGACGAGATACATGCTGGAAGGAGGCTTCATGCTGAGGCGGCGGACAG 322
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                QY 70 GCTGTGCGGTATGATAGAGCGCTGTTGTAGAAACGCTGACCCACACAG 122
                1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                Db 323 ACATCTCTCTCGACGACGCGGCTGTATGCGGAGATGTGACACACGACGCTG 375
                1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                RESULT 5
                CDS0037Q 1101 bp cna linear GSS U3-JUN-1999
                LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
                DEFINITION BACR08K14 of RPCI-98 library from Drosophila melanogaster (fruit
                fly), genomic survey sequence.
                ACCESSION AL064465.1 GI:4941722
                VERSION AL064465
                KEYWORDS GSS.
                ORGANISM Drosophila melanogaster.
                Drosophila melanogaster.
                Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
                Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                Ephydroidea; Drosophilidae; Drosophila.

```

```

REFERENCE 1 (bases 1 to 1101);
GENSCOPE Genoscope.
AUTHORS Direct Submission:
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL SP 191 91006 EVRY cdex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoeawa and
            Aaron Mammos in Pletor de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA and provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
            Location/Qualifiers
                1..1101
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone="BACR08K14"
                /clone_lib="RPCI-98"
                /note="end : TET3"
            BASE COUNT 374 a 61 c 78 g 129 t 459 others
            ORIGIN
                Query Match 2.5%; Score 36.2; DB 17; Length 1101;
                Best Local Similarity 12.2%; Pred. No. 11;
                Matches 49; Conservative 173; Mismatches 181; Indels 0; Gaps 0;
                QY 186 GAGATAACATATGAAATGGTTGGAAGAATCCATTATGGCCAAACGCGGTGGTGGTGGC 245
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 593 KRGVMBDNVKAATDCTKTKMAVAMARAGAAAMARTAMRAAKAAKAKKKKGTR 752
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                QY 246 GCGTAACCGGTACCGATCACTGACGGAAGAATGCAAAAGAGATTTCATACACCA 315
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 752 ATCKKMKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 812
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                QY 306 TGGCGCTTATTCACACCGCGTCTGACCATCGAAGCGGTGACCGATTATTCGCACA 365
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 613 KKGAMVVKAAKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 872
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                QY 346 CTCGAGATGCTTTGAGAGTGCTATCAATTCGAGACAGGATATCCGAGCGAGTTGCTAA 425
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 575 VKSKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHKHK 932
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                QY 426 AATGCCCTTTCTCAACCCACAAACGCGGATCGTCAATGGCGCGGAGAAAGCTG 495
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 943 VNRVNVKVKVAVVMGVGMVVMVVMVVMVVMVVMVVMVVMVVMVVMVVMVVMVVM 992
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                QY 466 ATGTGCTCGCTGTCTATATCGAATCCATGTTCGCCCGCGCGCTTGATCCCTACGCGATCT 545
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 593 GVCVBNKKKNGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNG 1052
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                QY 546 CGCCCATGATTCGCGATTTTGGCGGACGTGACCGCGGACGACCT 588
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                Db 1053 NMKGHVTRDWTGRRGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGK 1095
                111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
                RESULT 6
                AV629284 495 bp mRNA linear EST 15-DEC 2000
                LOCUS AV629284 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
                DEFINITION reinhardtii cDNA clone LCL055H05_r 5', mRNA sequence.
                ACCESSION AV629284
                VERSION AV629284.1 GI:10791918
                KEYWORDS EST
                SOURCE Chlamydomonas reinhardtii.

```

Contact: Blaxter M  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JT, UK  
Tel: +44 131 650 8760  
Fax: +44 131 576 3450  
Email: mark.blaxter@ed.ac.uk

Sequenced from the Brugia malayi BAC library constructed by Claire Whitten and Dr Mike Quail. The sequence was generated by the Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICAPB, University of Edinburgh, Edinburgh, UK.

Seq primer: SP6 (ATTAGGTGACACTATAG)  
Class: BAC ends.  
Location/Qualifiers  
1 . 482  
    /organism="Brugia malayi;"  
    /strain="TRS"  
    /db\_xref="taxon:6279"  
    /cdate\_lib="Brugia malayi Genomic Bac Library 3"  
    /sex="Mixed (male and female)"  
    /tissue\_type="whole parasite"  
    /dev\_stage="microfilaria (L1)"  
    /noc="vector: pHACE3.6; Site\_1: BamH I; Bug-a-mayo;  
genomic DNA was partially cleaved with Sau3A 2 and size fractionated. 7,392 clones were generated with mean insert size ~48 kbp. The library was constructed by Claire Whitten, Blaxter Nematode Genetics Lab, University of Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, The Sanger Centre, Cambridge, UK."  
BASE COUNT     166 a   89 c   45 g   178 t   1 others  
ORIGIN

Query Match.                 2.5%   Score 35.6; DB 17; Length 482;  
Best Local Similarity     54.2%; Pred. NO. 9.5;  
Matches     71; Conservative     0; Mismatches     60; Indels     0; Gaps     0;

QY     72 TGTGGTGATGAATAAAGCCCTGGTTGTAGAACAACCGTCACCACCAACAGCCTCTGCAT 131  
        ||| |||| | || | ||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB     236 TGTTIGTATITGTAATTCCTCATTTGAAGACACAGATGAACAAGATAAGCTTTCAGTT 227  
  
QY     132 GATCTTTTATGCCCTGCATCTATGGCTCTGTACCTAAACGGTATAAATTAGCTGGAGAAAT 191  
        ||| ||| | || | || | || | || | || | || | || | || | || | || | || |  
DB     236 GAICGTATTATGACAGCAATGATTTTGATATGAANGGAATAATATTTTTTTCAGAAA 167  
  
QY     192 AACCATATGAAA 202  
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB     166 AAAAGGTAAAA 156  
  
RESULT #  
LOCUS AGI17547  
DEFINITION Pan troglodytes DNA, clone: PTB-125G18.F, genomic survey sequence.  
ACCESSION AGI17547  
VERSION AGI17547.1 GI:16738066  
KEYWORDS PAN  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB-Chimpanzee Male.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Eumetazoa; Mamalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.  
  
REFERENCE  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki.Y.  
TITLE RAC end sequences of Library PTB unpublished  
JOURNAL  
REFERENCE 2 (bases 1 to 735)  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki.Y.  
DIRECT SUBMISSION  
TITLE Submitted (02-AUG-2001) Asao,Fujiyama, The Institute of Physiol  
JOURNAL

and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suehiro-chou, Tsukuba-shi, Ibaraki, Japan  
(E-mail: chimbase@sc.riken.go.jp, URL: http://hjb.sc.riken.go.jp/  
tel: 81-45-503-9111, fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library FIB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 735

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-125G18.F"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

188 a 186 c 160 g 198 t 3 others

BASE COUNT

ORIGIN

Query Match 2.58; Score 35.6; DB 17; Length 755;

Best Local Similarity 58.5%; Pred. No. 13;

Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1315 TACTCTTCAGCGAGTGGCGGCGGAGAGTGTCCACATGGCGGGGAGCCGAG 1374

DB 123 TGCCTTCAGCGGAGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAG 142

QY 1375 GCTATTCGGAATTAATCAATATGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAG 1422

DE 163 GTTGTGACCAATTAATTAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 228

RESULT 9

BE442656

LOCUS

DEFINITION

WHE1101\_D03\_G052S Wheat etiolated seedling root normalized cDNA

library Triticum aestivum cDNA clone WHE1101\_D03\_G05, mRNA

sequence.

ACCESSION

BE442656

VERSION

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum

bread wheat.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

Triticeae; Triticum.

REFERENCE

1 (bases 1 to 611)

Anderson, O.D., Chao, S., Choi, E.W., Close, T.J., Fiedler, R.D., Han

P.S., Hsiao, C.C., Kang, Y., Lazo, G.R., Miller, S., Nguyen, H.T.,

Rausch, C.C., Seaton, C.L., Tong, J.C., and Zhang, D.

The structure and function of the expressed portion of the wheat

genomes - Normalized root cDNA library

Unpublished (2000)

JOURNAL

COMMENT

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: andersnew.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratiagene SK primer.

Location/Qualifiers

1. 611

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

FEATURES

Source

1. 772

/organism="Brassica oleracea"

/strain="T01000DH3"

/clone="WHE1101\_D03\_G05"

/clone\_lib="Wheat etiolated seedling root normalized cDNA

library"

/tissue\_type="Root"

/dev\_stage="Five day old etiolated seedling"

/lab\_host="E. coli DH10B"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid

pBluescript SK; Site 1: EcoRI; Site 2: XhoI; Seeds were

surface-sterilized, germinated and grown aseptically in

the dark at room temperature on filter paper with water,

nystatin and cefotaxime in covered crystallization

disks. Roots were harvested. The tissue, total RNA, and

poly(A) RNA were prepared. A cDNA library was made in the

TJ Close lab (Choi, Close, Fenton) at the University of

California, Riverside. The cDNA clones were in vivo

excised to give pBluescript phagemids before

normalization was carried out. The mass excision of

phagemid library and normalization were done in HT Nguyen

lab by D. Zhang at Texas Tech University. Normalization

protocol used was that of Soares. Plasmid DNA

preparations and DNA sequencing were performed in the OD

Anderson lab (all other authors)."

BASE COUNT 181 a 111 c 153 q 166 t

ORIGIN

Query Match 2.48; Score 35.2; DB 10; Length 611;

Best Local Similarity 51.2%; Pred. No. 15;

Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 719 GAATTTGACTCAATCAATTCACAGCCGAGACATCAAGCGGGGAGTATGGAATGGCGG 778

DB 141 GCNATGAGCAGCAGCAATATTCCTGGAGAGCAATGGGGGAGATTCGACATTAAG 200

QY 779 GATATAGCAGCAGGAGTATTACATATCTGCGGTAGCTGCGGCTGGAGCGCCCTGT11 848

DB 231 AATCTAGCAGAGGTTTCAAGTTCATCTACCATTTTGTGTAAGGAGCAATCTGAGT 260

QY 839 ATTGTGATGCCCATGTTGTCAGGTCATGTCAGATT 878

DB 261 ACTGCGAATGCACTTTCCACAGGCGATGGTGAAGTCT 300

RESULT 10

BE769662/C

LOCUS

DEFINITION

WHE1071F B0\_2\_1\_K3 Brassica oleracea genomic clone B0MEL07, DNA

sequence.

ACCESSION

BE769662

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

1 (bases 1 to 772)

Town, C.D., Van Aken, S., Lutterback, T., and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

JOURNAL

COMMENT

Contact: Chris Town

7712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: 1P

Class: sheared ends.

Location/Qualifiers

1. 772

/organism="Brassica oleracea"

/strain="T01000DH3"



```

/sex="male"
/dev_stage="23 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAGAATTAATTAAGATCTTTTITTTTITTTT 3']. double-stranded cDNA was ligated to Eco RI adapters (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      89 a  104 c  107 g  113 t      # others
ORIGIN
Query Match      2.4%  Score 35;  DB 14;  Length 421;
Best Local Similarity 51.6%  Pred. No. 13;
Matches 80;  Conservative 0;  Mismatches 75;  Indels 0;  Gaps 0;

CY 1044 ACTGGCTGGTAGAAGACTTTGGCTTCGACAAATGGGATGCTTCTTCAGTCAAT 1103
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
DB 242 ACTGGGAAGAGATGACTTTGGCTGCGAGGATGATCTCAACGCGAGTCTGAGACAT 183
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
CY 1104 CGGCAAAAGTCGGCTGGGCAACATGCTGACCCCAATACACGCTGGCGGATCTGA 1163
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
DB 182 CCCACAGGAGTCTGCTGGGGGAGATGCCGTGACAGTTACACCGTGGGGTGTAACC 123
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
CY 1164 ACAAAACCTGTTAGTTTGTAGTAGGAATAAATACCC 1198
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
DB 122 ACTATACCTTATAATAGAGCTGATCTTTCTACCC 88
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||

RESULT 13
AA019639/c 441 bp  mRNA  human  EST 10-JAN-1997
LOCUS
DEFINITION
IMAGE:363600 3' similar to WP:K02A2.3 CE02791: SOMETANILU-SENSITIVE
NA-K-Cl cotransporter ; mRNA sequence.
AA019639 1 31:1483067
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 441)
Chissole, S., Dietrich, K., DuBuque, T., Pavello, A., Gish, W., Hawkins,
M., Huitman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Margis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, R.,
Underwood, K., Wohlmann, P., Waterston, K., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044479
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2010 Std Error: 0.00
Seq primer: 40M13 fwd. from Amersham
High quality sequence stop: 360.
Location/Qualifiers
1..441
/organism="Homo sapiens"
/db_xref="CDB:1280303"
/db_xref="Eaxon:9606"
/clone="IMAGE:363600"
/clone_lib="Soares retina N2b4HR"

FEATURES
source

```

```

/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: p773D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGGCGCGCTTTTITTTTITTTT 3']. double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)- RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      65 a  137 c  120 g  113 t      5 others
ORIGIN
Query Match      2.4%  Score 35;  DB 9;  Length 441;
Best Local Similarity 50.6%  Pred. No. 14;
Matches 83;  Conservative 0;  Mismatches 81;  Indels 0;  Gaps 0;

CY 535 CTACGGATCTGCCCATGATTCGCGATTTTGGCGGACTGACCGGACCGACCTGACGCG 594
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
DB 167 CAAGGACTCTTTCATCATGAAGCGGAGTGGGAGAACTTGAACCACTCCGACGTGGCGC 128
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
CY 595 CATGCTCAATGTCGCTGCCGAAAGGTCGCGATGATTAACTGCACTGAAAAGGT 654
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
DB 107 CATGCACTGCTGCGCTCAAGGAGTCTGTCGAGAAATCCCGGACGCTCAAGCT 48
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
CY 655 CTACTGAGCAAGACCAATAGCTTCCTATTAACCCCATATTG 698
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||
DB 47 TCTTTGCTCAACATGCTGGGCTTCCCGCAACGCAATGCTG 4
      |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||

RESULT 14
AA001557 492 bp  mRNA  linear  EST 29-NOV-1996
LOCUS
DEFINITION
IMAGE:362027 3' similar to WP:K02A2.3 CE02791: EUMFANILU-SENSITIVE
NA-K-Cl cotransporter ; contains Alu repetitive element.; mRNA
Sequence.
AA001557 1 31:1437022
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 492)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
M., Huitman, M., Kucaba, T., Le, M., Lennon, S., Maria, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, R.,
Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 3433 Std Error: 0.00
Seq primer: mob.REGA-ET
High quality sequence stop: 233.
Location/Qualifiers
1..492
/organism="Homo sapiens"
/db_xref="GDB:1278730"

```



```

..505
Source
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC IPiG"
/tissue_type="pooled"
/lab_host="DH10P"
/nc-c-vector="pCMV SPORT6; Site_1: NofL; Site_2: Sail;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT      104 a      121 c      158 g      122 t
ORIGIN
      2-44: Score 35; DB 12; Length 505;
      Best Local Similarity 55.33; Pred.No. 15;
      Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0
QY      2  CGCGGAAGCTCGCATGTGGCCGTGATCCGTGTCGACGAGGATATTCGATGATCCAGCGGAC 61
Cb      134  CCTCAATCTTCACCTGGGTCTCAGTTCGTGGTCCCTCAAGTGCAATCGCGTCTCTTGGGAT 393
QY      62  CCACAGAGCGCTGTGGGTATGATAAGGCGTGGTGTGTAGAACCGCTGACCCACACACA 121
Cb      194  TGAACTAGCGTGTAGGACCTGACCGTGGCGCTGACGCGGAGGAGCTGACGCTCCACCA 453
QY      122  GCT 124
Cb      454  GCT 456
Search completed: July 6, 2003, 05:40:28
Job time : 1978 secs

```

```

/db_xref="taxon:9606"
/clone="INAC:362027"
/clone_lib="Soares retina N244K"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/notes="Organ: eye; Vector: pT730 (Pharmacia); with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not 1 - oligo(dT) primer; 15' TGTTACCACTCAAGTCGACGCGCGGCTTTTCTTTTCTTTT 3']; double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not 1 and cloned into the Not 1 and Eco RI sites of a modified pT730 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldi."
BASE COUNT      80 a 141 c 131 g 129 t 11 others
ORIGIN
Query Match      2.4% Score 351 55 91 Length 422
Best Local Similarity 50.6%; P-adj. No. 15;
Matches 84: Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 535 CTACGGCATCGCGCATGATTCGGCATTTTGGCGACATGACCGGAGACGACGACGACG 534
166 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
167 CAGGAGCTCTTCAGCATGAGCGGAGTGGGAGAGACTTGAACCAAGTCCCAACGTCCGCGG 168
QY 535 CATGCTCAATGATCGCGTGCAGAAAGGTGGCGATGATTAAATCGACATGAAAGGT 534
166 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 CATGCACATGCTCGCGCTGAACGAGGATCGTGGAGGAATCCCGGAGCAATGAT 48
QY 535 CTATGTGAGCAAGCCGCAACGTTCCGTATTAACGCTATATG 598
166 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
167 TCTTTCTCAACATCGCTGGCGCTCCCGCCGACACGCAATGATG 4
RESULT 15
RG334211
TAXUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Simia; Suidae; Sus.
1 (bases 1 to 505)
Fairenkrug,S.C., Preking,B.A., Kohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68903-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980504.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCATGACCAT
BACKWARD: GTTITCCAGTCACGACG
Plate: 113 Row: 1 Column: 2
Seq primer: ATTAGTGACATATAS.
Location/Qualifiers

```

GenCore version 5.1.6  
Copyright (c) 1995 - 2003 CompuGen Inc.

CM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 04:47:45 ; Search time 97 seconds  
(without alignments)  
4559.050 Million cell updates/sec

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 cccgggaactccatg:ggcc.....aatgcaattcttgcaccc 1442

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum US seq length: 6

Maximum DB seq length: 2000000200

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37.5	2.6	503	1	US-08-123-343A-15
C 2	37.5	2.6	1233	4	US-09-431-573-1
C 3	37.5	2.6	1233	4	US-09-431-573-2
C 4	37.6	2.6	1233	4	US-09-431-573-3
C 5	37.6	2.6	3909	1	US-08-123-343A-6
C 6	35.2	2.4	429	4	US-09-144-367-7
C 7	34.2	2.4	5733	2	US-08-473-533A-1
C 8	33.2	2.3	3201	2	US-08-633-770A-1
C 9	33.2	2.3	4726	1	US-08-633-770A-2
C 10	32.8	2.3	7218	1	US-08-232-453-14
C 11	32.4	2.2	3176	1	US-09-305-572-3
C 12	32.4	2.2	9551	4	US-08-056-260-23
C 13	32.4	2.2	9551	2	US-08-800-644-53
C 14	32.4	2.2	1226	1	US-08-374-656-4
C 15	32.2	2.2	1349	1	US-08-374-656-1
C 16	32.2	2.2	6254	4	US-08-921-2-9-126
C 17	31.8	2.2	1917	4	US-08-943-731-193
C 18	31.8	2.2	20084	4	US-08-943-731-5
C 19	31.6	2.2	1380	4	US-09-513-783A-169
C 20	31.2	2.2	2167	3	US-08-743-637B-16
C 21	31.2	2.2	2167	3	US-08-525-840B-16
C 22	31	2.1	4403765	4	US-09-103-840A-2
C 23	31	2.1	4411529	4	US-09-103-840A-1
C 24	30.8	2.1	1187	5	PCT-US95-13536-2
C 25	30.8	2.1	1188	5	PCT-US95-13536-1
C 26	30.6	2.1	1542	4	US-09-202-893B-3
C 27	30.6	2.1	4403765	4	US-09-103-840A-2

C 28	30.6	2.1	4411529	4	US-09-103-840A-1
C 29	30.4	2.1	1049	3	US-08-652-643-1
C 30	30.2	2.1	747	4	US-08-958-416-1016
C 31	30.2	2.1	10095	3	US-08-822-586-45
C 32	30	2.1	1614	4	US-09-522-217-7
C 33	30	2.1	2615	3	US-09-840-905-1
C 34	29.8	2.1	289	4	US-09-007-005-17
C 35	29.8	2.1	289	4	US-09-244-796-17
C 36	29.8	2.1	417	4	US-08-936-165A-17
C 37	29.4	2.0	1588	4	US-08-485-107-1
C 38	29.4	2.0	7952	4	US-08-961-527-112
C 39	29.2	2.0	1524	4	US-09-134-031C-2055
C 40	29.2	2.0	45899	1	US-08-471-119A-1
C 41	29	2.0	300	1	US-08-083-946-6
C 42	29	2.0	300	3	US-08-452-915-6
C 43	29	2.0	615	4	US-09-134-001C-138
C 44	29	2.0	850	1	US-07-611-528A-1
C 45	29	2.0	850	1	US-08-083-946-1

#### ALIGNMENTS

RESULT 1  
US-08-123-343A-15/c  
Sequence 15, Application: US/98123343A  
Patent No. 593879  
GENERAL INFORMATION:  
APPLICANT: Steller, Hermann  
APPLICANT: Adams, John M.  
APPLICANT: Grether, Megan E.  
APPLICANT: White, Kristin  
TITLE OF INVENTION: Cell Death Genes of Drosophila  
TITLE OF INVENTION: Melanogaster and Vertebrate Analogs  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,343A  
FILING DATE: 17-SEP-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/104,957  
FILING DATE: 15-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Granchan, Patricia  
REGISTRATION NUMBER: 52,227  
REFERENCE/DOCKET NUMBER: MIT-5907A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-651-6240  
TELEFAX: 6186-9340  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-123-343A-15

Query Match 2.6% Score 37.6; DB 1; Length 503;  
Best Local Similarity 48.6%; Pred. No. 0.02; 109; Indels 0;  
Matches 103; Conservative 0; Mismatches 0; Gaps 0;

QY 320 ACACCGCTCCTGACCATCGAACCCGGTGACCGGATTATTGCGACACTCGAGATGGTITT 379  
DB 334 ACCACCGGACTGGCTGCGGTATGCGACACTGGATGATGCTGCTGCGAGTGGCTATT 275  
QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCGAGTTGCTTAAATGCGCTTCTC 439  
DB 274 GAGGTCTCTTCTTCGCGGCGGATCGCTCCATTAACCTCTCGACACTCGCGGTCG 215  
QY 440 AACCCACAAAACGCGGATCGTCAATTCGCGCGGAGAAAGGTGATGCTGCGCTTC 499  
DB 214 GCGCGTAAGTTGCTAGGATCGCCAAACTCGTCCCAATGCGCTCAATGCGCTGCTG 155  
QY 500 TAATCGATTCATTTGCGCGCGGCGGTGA 531  
DB 154 GTACTCGCGCTCATCTCGCGCGGTGAGTCA 123

RESULT 2  
US-09-431-573-1/c  
: Sequence 1, Application US/09431573  
: Patent No. 6235524  
: GENERAL INFORMATION:  
: APPLICANT: STELLER, HERMAN  
: APPLICANT: AGAPITE, JULIE  
: APPLICANT: MC CALL, KIMBERLY  
: APPLICANT: BERGMANN, ANDREAS  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING AGENTS THAT  
: TITLE OF INVENTION: INHIBIT MAPK MEDIATED ANTI-APOPTOTIC SIGNALS  
: FILE REFERENCE: MIT-04019  
: CURRENT APPLICATION NUMBER: US/09/431,573  
: PRIOR FILING DATE: 1999-10-29  
: PRIOR APPLICATION NUMBER: 60/106,108  
: NUMBER OF SEQ ID NOS: 5  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 1  
: LENGTH: 1233  
: TYPE: DNA  
: ORGANISM: Drosophila melanogaster  
US-09-431-573-1

Query Match 2.6%; Score 37.6; DB 4; Length 1233;  
Best Local Similarity 48.6%; Pred. No. 0.035;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
QY 320 ACACCGCTCCTGACCATCGAACCCGGTGACCGGATTATTGCGACACTCGAGATGGTITT 379  
DB 1059 ACCACCGGACTGGCTGCGGTATGCGACACTGGATGATGCTGCTGCGAGTGGCTATT 1000  
QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCGAGTTGCTTAAATGCGCTTCTC 439  
DB 999 GGAGCTCTTCTTCGCGGCGGATCGTCCATTAACCTCTCGACACTCGCGGTCG 940  
QY 440 AACCCACAAAACGCGGATCGTCAATTCGCGCGGAGAAAGGTGATGCTGCGCTTC 499  
DB 939 GCGCGTAAGTTGCTAGGATCGCCAAACTCGTCCCAATGCGCTCAATGCGCTGCTG 860  
QY 500 TAATCGATTCATTTGCGCGCGGCGGTGA 531  
DB 479 GTACTCGCGCTCATCTCGCGCGGTGAGTCA 848

RESULT 3  
US-09-431-573-2/c  
: Sequence 2, Application US/09431573  
: Patent No. 6235524  
: GENERAL INFORMATION:  
: APPLICANT: STELLER, HERMAN  
: APPLICANT: AGAPITE, JULIE  
: APPLICANT: MC CALL, KIMBERLY  
: APPLICANT: BERGMANN, ANDREAS  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING AGENTS THAT

: TITLE OF INVENTION: INHIBIT MAPK MEDIATED ANTI-APOPTOTIC SIGNALS  
: FILE REFERENCE: MIT-04019  
: CURRENT APPLICATION NUMBER: US/09/431,573  
: PRIOR FILING DATE: 1999-10-29  
: PRIOR APPLICATION NUMBER: 60/106,108  
: PRIOR FILING DATE: 1999-10-29  
: NUMBER OF SEQ ID NOS: 5  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 2  
: LENGTH: 1233  
: TYPE: DNA  
: ORGANISM: Drosophila melanogaster  
US-09-431-573-2

Query Match 2.6%; Score 37.6; DB 4; Length 1233;  
Best Local Similarity 48.6%; Pred. No. 0.035;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
QY 320 ACACCGCTCCTGACCATCGAACCCGGTGACCGGATTATTGCGACACTCGAGATGGTITT 379  
DB 1059 ACCACCGGACTGGCTGCGGTATGCGACACTGGATGATGCTGCTGCGAGTGGCTATT 1000  
QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCGAGTTGCTTAAATGCGCTTCTC 439  
DB 999 GGAGCTCTTCTTCGCGGCGGATCGTCCATTAACCTCTCGACACTCGCGGTCG 940  
QY 440 AACCCACAAAACGCGGATCGTCAATTCGCGCGGAGAAAGGTGATGCTGCGCTTC 499  
DB 939 GCGCGTAAGTTGCTAGGATCGCCAAACTCGTCCCAATGCGCTCAATGCGCTGCTG 860  
QY 500 TAATCGATTCATTTGCGCGCGGCGGTGA 531  
DB 479 GTACTCGCGCTCATCTCGCGCGGTGAGTCA 848

RESULT 4  
US-09-431-573-3/c  
: Sequence 3, Application US/09431573  
: Patent No. 6235524  
: GENERAL INFORMATION:  
: APPLICANT: STELLER, HERMAN  
: APPLICANT: AGAPITE, JULIE  
: APPLICANT: MC CALL, KIMBERLY  
: APPLICANT: BERGMANN, ANDREAS  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING AGENTS THAT  
: TITLE OF INVENTION: INHIBIT MAPK MEDIATED ANTI-APOPTOTIC SIGNALS  
: FILE REFERENCE: MIT-04019  
: CURRENT APPLICATION NUMBER: US/09/431,573  
: PRIOR FILING DATE: 1999-10-29  
: PRIOR APPLICATION NUMBER: 60/106,108  
: NUMBER OF SEQ ID NOS: 5  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 3  
: LENGTH: 1233  
: TYPE: DNA  
: ORGANISM: Drosophila melanogaster  
US-09-431-573-3

Query Match 2.6%; Score 37.6; DB 4; Length 1233;  
Best Local Similarity 48.6%; Pred. No. 0.035;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
QY 320 ACACCGCTCCTGACCATCGAACCCGGTGACCGGATTATTGCGACACTCGAGATGGTITT 379  
DB 1059 ACCACCGGACTGGCTGCGGTATGCGACACTGGATGATGCTGCTGCGAGTGGCTATT 1000  
QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCGAGTTGCTTAAATGCGCTTCTC 439  
DB 999 GGAGCTCTTCTTCGCGGCGGATCGTCCATTAACCTCTCGACACTCGCGGTCG 940  
QY 440 AACCCACAAAACGCGGATCGTCAATTCGCGCGGAGAAAGGTGATGCTGCGCTTC 499  
DB 939 GCGCGTAAGTTGCTAGGATCGCCAAACTCGTCCCAATGCGCTCAATGCGCTGCTG 860

```
339 GCGCGTAAAGTTGCTAGGATCGGCAAACTGTCUCAAAGTGGGCTCAAGTCGGTCTG 880
500 TATATCGAATCCATCTTGGCCCGGCGGCTTGA 531
111 111 111 111 111
879 GTACTCGGCTCATCTCGGCGGCTCGAGGTCA 848

RESULT 5
US-08-123-443A-6/c
: Sequence 6, Application US/0812343A
: Patent No. 5593679
: GENERAL INFORMATION:
: APPLICANT: Steller, Hermann
: APPLICANT: Abrams, John M.
: APPLICANT: Grether, Megan E.
: APPLICANT: White, Kristin
: TITLE OF INVENTION: Cell Death Genes of Drosophila
: TITLE OF INVENTION: Melanogaster and Vertebrate Analogs
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: US
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08/123.343A
: APPLICATION NUMBER: US/08/123.343A
: FILING DATE: 17-SEP-1993
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/004.957
: FILING DATE: 15-JAN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: MIT-5907A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 61861-9540
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3900 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
US-08-123-343A-6

Query Match 2.6% Score 37.6; DB 1; Length 3900;
Best Local Similarity 48.6% Pred. No. 0.072;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 320 ACACCGTCTGACCATCGGACCGGTGACCGGATTATTTCGACATCGGATGCTT 379
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 1458 ACCACCGGACTGCTGCGGTATGGCAGACGGGATTATTCGTCGAGTGACAT 1399

QY 380 GAAGTGCTATCATTCGGACACAGCATTCGAGGCGAGTTGCTAATAATCGCTTCT 439
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 1398 GGAGCTCTTCTTCGCGGCGGGATGGCTCCATTGCAACTCCGACAGGCTCCGCG 1339

QY 440 AACCCACAAACGACCGATCATGCTCAATGGCGGAGAAAGTGAIGTCTCGCTG 499
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 1338 GCGCGTAAGTTGTGTAGGATGCCCAAACTGCTCCCAAGTGGCTCATGATCGCTC 1279

QY 500 TATATCGAATCCATGTTGCCCGGCGGCTTGA 531
111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 1278 GTACTCGGCTCATCTCTCGCGGCTCGAGGTCA 1247
```

```
RESULT 6
US-09-144-367-2/c
: Sequence 7, Application US/09144367
: Patent No. 642659
: GENERAL INFORMATION:
: APPLICANT: Lichteir, Jay
: APPLICANT: Guido, Marco
: TITLE OF INVENTION: GEN-TYPING OF HUMAN CYB3A4
: FILE REFERENCE: SEQ-12P
: CURRENT APPLICATION NUMBER: US/09/144.367
: CURRENT FILING DATE: 1998-08-31
: PRIOR APPLICATION NUMBER: 60/058.612
: PRIOR FILING DATE: 1997-09-10
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 7
: LENGTH: 429
: TYPE: DNA
: ORGANISM: H. sapiens
: FEATURE:
: NAME/KEY: Other
: LOCATION: (0)...(0)
US-09-144-367-7

Query Match 2.4% Score 35.2; DB 4; Length 429;
Best Local Similarity 57.5% Pred. No. 0.12;
Matches 61; Conservative 1; Mismatches 44; Indels 0; Gaps 0;

QY 1164 ACAAAAACGCTAGTCTAGTGGTAACTAACCGGTGACATTAACCGGATGATG 1223
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 315 ACACATATCTTCAATGCTACTACAACTGCTGAACTGATATTTAAGTGGATGAT 256
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 1224 GGGTAATGTGAAGTTCACAACTCGCTATTTTAAACAGCTAAAG 1269
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 255 ATGGTGATTATCTCAATAAGGAGTTATTTTAAGAGACGAAAG 210
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

RESULT 7
US-08-473-553A-1
: Sequence 1, Application US/08473553A
: Patent No. 5859388
: GENERAL INFORMATION:
: APPLICANT: Moxrowitz, Elliot M.
: APPLICANT: Clark, Steven E.
: APPLICANT: Williams, Robert W.
: TITLE OF INVENTION: Plant Clavatal Nucleic Acids,
: TITLE OF INVENTION: Transformed Plants, and Proteins
: NUMBER OF SEQUENCES: 1;
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hebeach, Test, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3450
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/473.553A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REGISTRATION NUMBER: 38,304
: REFERENCE/DOCKET NUMBER: A-60886/RET/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
```

TELEX: 510 277299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5733 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:

NAME/KEY: CDS  
LOCATION: 2434..5037  
FEATURE: CDS  
NAME/KEY: CDS  
LOCATION: 5117..5467  
US-08-473-553A-1

Query Match 2.4%; Score 34.2; DB 2; Length 5733;  
Best Local Similarity 50.3%; Pred. No. 1.2;  
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1136 CCCAATACACCGTTCGGCGGATGCTGACAAAGAACCTGTTAGTTAGTACGAATACGA 1195  
DB 3709 CTCAGGGGACGTTCCGGGGGCTTTCAATCTACGTTTATACGATTACGATC 3766  
QY 1196 ACCGGTGAACATACCGGATGTAGATCGGGTAAATGTGTAATCAACAAATGCTAAT 1255  
DB 3769 ACTGATAATTCCTCTCCGATGACCTCCGTAACGATGTCGAGGATGTTCAATGATG 3828

QY 1256 TTACAGCTAAGACGTCGCATATGAGCGGACGATACGATACGATACGATACGATACG 1302  
DB 3829 ATTACCTCTCTACAAACGTTTTCGGGCGAGATCCACCTCCGAT 3878

## RESULT 8

US-08-633-770A-3/c  
Sequence 3, Application US/08633770A  
Patent No. 5908760

## GENERAL INFORMATION:

APPLICANT: Bojlsen, Kirsten  
APPLICANT: Yu, Shukun  
APPLICANT: Kragh, Karsten  
APPLICANT: Christensen, Torv  
APPLICANT: Marcussen, Jan  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,770A  
FILING DATE: July 8, 1996

## CLASSIFICATION: 435

PRIOR APPLICATION NUMBER: PCT/EP94/03398  
FILING DATE: OCT-15-1994

## ATTORNEY/AGENT INFORMATION:

NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY006.001APC

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3201 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-633-770A-3

Query Match 2.3%; Score 33.2; DB 2; Length 3201;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 CCGGGCGTAACCGGTAAACCGTAAACCGTAAACCGTAAACCGTAAACCGTAAACCGTAA 302  
DB 1228 CTGGGCTTCACTTTGAACGTCCTCAACGACCTGTATATAAGTCTGTACAAATTCGGGAC 969  
QY 303 CCATTGGCCCTTATTCACACCCGCTCTGACCATCGAACCCGCTGACCGGATATCTCG 362  
DB 968 GTATCCGACATGATACCGTAAACGATACGATACGATACGATACGATACGATACGATAC 909  
QY 363 ACATTCGAGATGCTTTTGAAGTCTCTATCAATTCGAGACGATAT 408  
DB 968 CTTGATGCTTCTTCAACGATCGATCGCAATTCGAGATGATAT 863

## RESULT 9

US-08-633-770A-11/c

Sequence 11, Application US/08633770A  
Patent No. 5908760

## GENERAL INFORMATION:

APPLICANT: Bojlsen, Kirsten  
APPLICANT: Yu, Shukun  
APPLICANT: Kragh, Karsten  
APPLICANT: Christensen, Torv  
APPLICANT: Marcussen, Jan  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,770A  
FILING DATE: July 8, 1996

## CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP94/03398  
FILING DATE: OCT-15-1994

## ATTORNEY/AGENT INFORMATION:

NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY006.001APC

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:

## INFORMATION FOR SEQ ID NO: 11:

## SEQUENCE CHARACTERISTICS:

LENGTH: 4726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

US-08-633-770A-11

Query Match 2.3%; Score 33.2; DB 1; Length 4726;  
Best Local Similarity 50.0%; Pred. No. 2.4; Mismatches 0; Gaps 0;  
Matches 83; Conservative 0; Indels 83; Gaps 0;  
CY 243 CCGGGGTAACCGGTAAAGCATCCTGACGAGAGAAATGCAAAAGAGTTTCAATACA 302  
DB 1741 CTGGGCTCAACTTTGAAGGTCACCAAGACCGTATAAAGTCGTACAAATTCGGGAGCC 1683  
CY 303 CCATTGGCCCTTATTCACACCGGCTCTGACCATCAACCGGTACCGGATATATGTCG 362  
DB 1681 GTATCGCACTGATACCGTAACANTCGATACCAACCATATCTGGTTCGACGTTGATAG 1622  
CY 363 ACATCGAGATGCTTTGAAGGTGCTATCAATTCGGAAAGGAGAT 438  
DB 1621 CCTGAGTGGCTCTTCAGAGTCGATGGCAATTTGAGAGTAGAT 1576

RESULT 10  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOMLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US/07/945.413  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.4  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,468  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 47031636-9300  
; TELEFAX: 893149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 728 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F15  
US-08-232-463-14

Query Match 2.3%; Score 32.8; DB 1; Length 7216;  
Best Local Similarity 3.6%; Pred. No. 4.2;  
Matches 13; Conservative 193; Mismatches 160; Indels 0; Gaps 0;  
CY 1051 GGTAGAGAGACTTTGGCTTCGAGCAATGGGATGCTACATGCTTCTGAGTCAATTCGGGCA 1110

DB 1452 CATAGAAGAAATTTGGTACRR 1393  
CY 1111 ACTCGCGCTGGSCAACAATGGTCGACCCCAATACACCGTTGGCGCGCATCTGACACAA 1170  
DB 1192 RRR 1353  
CY 1171 CCTTTAGTTTACTAGTAATAATACGCTGAACATTACCGGATGATAGTGGGTAA 1230  
DB 1342 RRR 1273  
CY 1251 TGTGTAGTTCACACATCSCATTTTAAACNACIANAGCGGTGATATGCGGCACAT 1290  
DB 1272 RRR 1213  
CY 1291 ACACCATCAATATGCTTTACTTTACTTCCTTCAGCGGAGTACGCGGSCACAAAGTTG 1350  
DB 1212 RRR 1153  
CY 1351 TCACAAATGGCGGAGCAACCCAGGUTATTCGGAATTAATCAAAATGGCGCATCAAC 1410  
DB 1152 RRR 1093  
CY 1411 GCCAGA 1416  
DB 1092 RRRRRR 1087  
RESULT 11  
US-09-309-572-3/c  
; Sequence 3, Application US/09309572  
; Patent No. 6440730  
; GENERAL INFORMATION:  
; APPLICANT: Heinrich-Pette-Institut  
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with HCMV  
; FILE REFERENCE: P50489  
; CURRENT APPLICATION NUMBER: US/09/309.572  
; CURRENT FILING DATE: 1999-05-11  
; EARLIER APPLICATION NUMBER: DE 198 56 463  
; EARLIER FILING DATE: 1998-11-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 3476  
; TYPE: DNA  
; ORGANISM: Lymphocytic choriomeningitis virus  
US-09-309-572-3  
Query Match 2.2%; Score 32.4; DB 4; Length 3376;  
Best Local Similarity 54.1%; Pred. No. 3.5;  
Matches 65; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
CY 426 AATGCGCTTTCTGACCCCAACAAACGACGATCATGTCATATGGCGGAGAAAGAGT 485  
DB 2141 AATAGCAATTTCCAAACACAGACGCGGAGTTTCACTACACTTTACAGGCAACGCTG 2082  
CY 486 ATGTGCTGCGTGCTATATCGAATCCATGTGGCCCGCGGTGATCGCTACGATCT 545  
DB 2701 ATCAAAACATTCAGACAGATTCGAGTACTACAGCGCATGCACTTCGCGAGCTCT 2022  
CY 546 GC 547  
DB 2021 TC 2020  
RESULT 12  
US-08-050-260-93  
; Sequence 93, Application US/08056200  
; Patent No. 5616500  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/04/056,220  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.0031A  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507...1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645...2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512...8070  
US-04-056-200-93

Query Match 2.2% Score 32.4; DB 1; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 3 CGGAACTCCATGCGCGTATGATCGAGCAGGATATTCGATGATCCAGCGGSC 62  
Db 3094 CGAGAGCGCAAGACAGAGTGTCCAGGAGGAGACAGAGTGCAGGAGCGCCAG 3123  
Qy 63 GCACAGCGCTGCGGTAATGGATAAGGCGTGGTTGTAGAACGCTGACCCANACAG 122  
Db 3124 ACAGTCTCCGGAAGGAAGAGAGTTTCAGAGGAAGAGAGCGCGGCAAGAGAG 3183  
Qy 123 CTCCTCATGATCTTTTAAATGCGTCTCTCTGCTCTGTAA 164  
Db 3184 CTCAGGAGGAAGGAAGAGAGCTACGAGAGCTGGAGCGGCAA 3225

RESULT 13  
US-04-056-200-93  
Sequence 93, Application US/04/0644  
Patent No. 5958752  
GENERAL INFORMATION:  
APPLICANT: Steiner, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: F-dppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507...1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645...2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512...8070  
US-08-800-644-93

Query Match 2.2% Score 32.4; DB 2; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 3 CGGAACTCCATGCGCGTATGATCGAGCAGGATATTCGATGATCCAGCGGSC 62  
Db 3064 CGAGAGCGCAAGACAGAGTGTTCAGGAGGAGAGAGAGTGCAGGAGCGCCAG 3123  
Qy 63 GCACAGCGCTGCGGTAATGGATAAGGCGTGGTTGTAGAACGCTGACCCANACAG 122  
Db 3124 ACAGTCTCCGGAAGGAAGAGAGTTCAGGAGGAGAGAGTGCAGGAGCGCCAG 3183  
Qy 123 CTCCTCATGATCTTTTAAATGCGTCTCTCTGCTCTGTAA 164  
Db 3184 CTCAGGAGGAAGGAAGAGAGCTACGAGAGCTGGAGCGGCAA 3225

RESULT 14  
US-08-800-644-93  
Sequence 4, Application US/08374686  
Patent No. 5616474  
GENERAL INFORMATION:





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 02:01:46 ; Search time 3753 seconds  
(without alignments)  
11182.053 Million cell updates/sec

Title: US-10-086-082-1  
Perfect score: 1442  
Sequence: 1 cccgggaactccatgtggcc.....aatgcaatcatttggatcc 1442

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455:402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pi.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1442	100.0	1442	6	A69475 Sequence 1
2	1442	100.0	1442	6	A72152 Sequence 1
3	74	5.1	339650	1	AP003583 Nostoc sp
4	61.4	4.3	13698	1	AE005929 Caulobact
5	60.4	4.2	360150	1	AP001507 Bacillus
6	59.6	4.1	10495	1	AE006875 Sulfolobu
7	58.8	4.1	7983	1	AE013142 Thermoplas
8	56	3.9	110030	2	LMFLCHR36_01
9	55.8	3.9	263050	1	AP000981 Sulfobolob
10	46.4	3.2	294800	1	SM5591789
11	45.6	3.2	167000	1	AP000059 Aeropyrum
12	42.8	3.0	35348	8	SPAC869
13	41.2	2.9	307120	1	CNSPAX03
14	41.2	2.9	349061	1	NMA222491
15	41.2	2.9	349980	6	AX041919 Neisseria
16	40.8	2.8	716	11	PM11HAG
17	40.8	2.8	793	11	PM9D96
18	39.6	2.7	92918	2	AC020050 Drosophill
19	39.6	2.7	130583	2	AC007420 Drosophill
20	39.6	2.7	155566	3	AC092231 Drosophill
21	39.6	2.7	170064	3	AC008323 Drosophill
22	39.6	2.7	304383	3	AE003579 Drosophill
23	39.4	2.7	118593	9	HS28811
24	39.2	2.7	1893	3	AY075188 Drosophill
25	39.2	2.7	10619	1	AF002546 Neisseria
26	39.2	2.7	133144	2	AC015424 Drosophill
27	39.2	2.7	150399	2	AC009369 Drosophill
28	39.2	2.7	183291	3	AC010003 Drosophill
29	39.2	2.7	183911	3	AC104507 Drosophill
30	39.2	2.7	271237	3	AE003521 Drosophill
31	39.2	2.7	349980	6	AX044034 Sequence
32	39.8	2.7	125620	9	AF429315 Homo sapi
33	38.4	2.7	10029	1	AE011694 Xanthomon
34	38.4	2.7	151085	8	AP003239 Oryza sat
35	37.6	2.6	503	6	I33795 Sequence 15
36	37.6	2.6	1233	6	AR153490 Sequence
37	37.6	2.6	1233	6	AR153491 Sequence
38	37.6	2.6	1233	6	AR153492 Sequence
39	37.6	2.6	3900	6	I33788 Sequence 6
40	37.6	2.6	3902	3	DMC31226 Drosophila
41	37.4	2.6	10029	1	AE009093 Agrobacte
42	37.4	2.6	14140	1	AE008058 Agrobacte
43	37.4	2.6	176688	2	AC084696 Mus muscu
44	37.4	2.6	162017	2	AL824710 Mus muscu
45	37.4	2.6	231853	2	AC093352 Mus muscu

ALIGNMENTS

RESULT 1  
A69475

LOCUS Sequence 1 from Patent WO9802551.  
DEFINITION A69475

ACCESSION A69475.1 GI:4760231  
VERSION

KEYWORDS

SOURCE

ORGANISM

Klebsiella oxytoca.

Klebsiella oxytoca

Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;

Klebsiella.

(bases 1 to 1442)

REFERENCE

AUTHORS Hovenberg, R.A., Van, D.L., Kerkman, R. and Nieboer, M.

TITLE IMPROVED PROCESS FOR THE PRODUCTION OF ADIPOYL CEPHALOSPORINS

JOURNAL Patent: WO 9802551-A 1 22-JAN-1998;

linear 1442 bp DNA PAT 06-MAY-1999

post dated

Pred. No. is the number of results predicted by chance to have a

> post dated.

```
FEATURES
Source
G:ST BROCADES BV (NL)
Location/Qualifiers
1..1442
/organism="Klebsiella oxytoca"
/strain="PRS1"
/isolate="PRS1"
/db_xref="taxon:571"
197..1183
/codon_start=1
/transl_table=1
/product="AMIDASE"
/protein_id="CAB42489.1"
/db_xref="GI:4760232"
/translation="MKWLESIMAKRGVAGRKPVTHLTETEMCKEFHYTIGPSTPV
LIIEPDGRIIVDRDAPEGAISEODIPOLKMPFINPCNPIHYNGAKGDLAVY
IESMLRGVDPYIGICAMIPHEGGLTGTDITAMLDNLPKVKRMIKLDSKYVWSRHT
LPYKHITLSVPEIDSSINSUTPDHNGMNOVPDIPGUSITYPVRAPGGRFLF-GGA
HACGGDEFGCTAVEFASITIKVDLIKMKQSLPMENAEINMSTGSAPIFIDATRI
AYRDLIYWLVED?GFQWDAYMILISQKVRILGNKVPYDPRYTVGAMENKNELV"
BASE COUNT 385 a 350 c 370 g 337 t
ORIGIN
Query Match 100.0%; Score 1442; DB 6; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGGGAAACCCATGTGGCGGTGATCTGCTGAGCAGGAGATATTGGGATGATCCAGCGG 60
DB 1 CCGGGAACTCCATGTGGCGGTGATCTGCTGAGCAGGAGATATTGGGATGATCCAGCGG 60
QY 51 CCGCAGACGGCTGTGGCGGTGATGATGAAGAAGCTGGTGTGAGAAAGCCTGACCCACAC 120
DB 61 CCGCAGACGGCTGTGGCGGTGATGATGAAGAAGCTGGTGTGAGAAAGCCTGACCCACAC 120
QY 121 AGCTCTGTGATGATCTTTAAAGCTGCTGATCTGCTGATCTGATCTGATCTGATCTGAT 180
DB 121 AGCTCTGTGATGATCTTTAAAGCTGCTGATCTGCTGATCTGATCTGATCTGATCTGAT 180
QY 181 AGCTGGAGATACATATGAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 AGCTGGAGATACATATGAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 TGGCGGCGGTAAACGGGTACGATACCTGACCGGAGAGAAATGCAAAAGAGTTCATTA 300
DB 241 TGGCGGCGGTAAACGGGTACGATACCTGACCGGAGAGAAATGCAAAAGAGTTCATTA 300
QY 301 CACCATTTGGCGCTTATTCACACCGCTGCTGACCATCGACCGGCTGACCGGATATGCT 360
DB 301 CACCATTTGGCGCTTATTCACACCGCTGCTGACCATCGACCGGCTGACCGGATATGCT 360
QY 361 CGACACTCGAGATGCTTTTGAAGTGTCTATCAATTCGGAACAGGATATTCGAGCGAGT 420
DB 361 CGACACTCGAGATGCTTTTGAAGTGTCTATCAATTCGGAACAGGATATTCGAGCGAGT 420
QY 421 GCTAAAAATGCCCTTTCTCAACCCACAAAACGACCGATCATGCTCAATGGCGCGAGAA 480
DB 421 GCTAAAAATGCCCTTTCTCAACCCACAAAACGACCGATCATGCTCAATGGCGCGAGAA 480
QY 481 AGGTGATGTGCTGCTGCTATATCGAATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 AGGTGATGTGCTGCTGCTATATCGAATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 CATCTGCGCCATGATTCGGATTTTGGCGGACTGACGGGACCGACCTGACGCGCATGCT 600
DB 541 CATCTGCGCCATGATTCGGATTTTGGCGGACTGACGGGACCGACCTGACGCGCATGCT 600
QY 601 CAATGATCCGCTGCCAAGAAAGTGGCGCATGATTAACCTCGACAGTGAAGAAGTCTACTG 660
DB 601 CAATGATCCGCTGCCAAGAAAGTGGCGCATGATTAACCTCGACAGTGAAGAAGTCTACTG 660
QY 661 GAGCAACGGCATACGCTTCCCTATACCCCGCATATGCGGACCTGTAGCGGATATCGGAGA 720
DB 661 GAGCAACGGCATACGCTTCCCTATACCCCGCATATGCGGACCTGTAGCGGATATCGGAGA 720
```

## RESULT 2

A72152

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

A72152 Sequence 1 from Patent WO9801568. 1442 bp DNA linear PAT 11-MAY-1999

A72152.1 GI:4808107

Klebsiella oxytoca.

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Klebsiella.

1 (bases 1 to 1442)

Brieden, W., Naughton, A., Robins, K., Shaw, N., Tinschert, A. and

Zimmermann, T.

METHOD OF PREPARING (S) - OR (R) - 3,3,3-TRIFLUORO-2-HYDROXY-2-

METHYLPROPIONIC ACID

Patent: WO 9801566-A 15-JAN-1998;

ONZA AG (CH)

Location/Qualifiers

applicant's own seq. 100% pure

source	1. .1442		
	/organism="Xiebiella oxyloca"		
	/strain="PRSI"		
	/isolate="PRSI"		
	/db_xref="taxon:57"		
CDS	197..1183		
	/codon_start=1		
	/transl_table=11		
	/product="AMIDASE"		
	/protein_id="CAB47722.1"		
	/db_xref="GI:480108"		
	/translation="MKWLESIMAKRGVCAKPKVTHHITSEMKKHHTIGVSTPV LEMPDRIVUTRADFEGALNSQDIPSQLKMPFLNPONGEIVNGAEGDLVAV IESMPLRGVDPYGCAMIPFGGLIGTDILAMLNDELPEKVRKIKLDSFKVYWSKRV LYPKHIGTISVSPIDNSLTIDNHHGNSXDPDIFGSIYLPKVRGKRLFIGDA HACQSGEIGCTAFVEFASITTKVDLIKWLSNPMEFNAENMISIGSRPLEDAF AYRDLIYMLVDFGFEOMDAYMLISQCKVRUGNKVTPKYTVGAMLNKLIV"		
BASE COUNT	385 a 350 c 370 g 337 t		
ORIGIN			
	Query Match: 100.0%; Score 1442; DB 6; Length: 1442;		
	Best Local Similarity 100.0%; Pred. No. 0;		
	Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	CCCGGAAGCTCCATGTGGCGGTGATCCTGGTGGACACAGATATTCGGGATGATCCAGCGG 60	
DB	1	CCCGGAAGCTCCATGTGGCGGTGATCCTGGTGGACACAGATATTCGGGATGATCCAGCGG 60	
QY	61	CCGCACAGCGTGTGGGTAAATGGATAAAGCGCTGCTGTGATAAAGCGTGCACCAACAC 120	
DB	61	CCGCACAGCGTGTGGGTAAATGGATAAAGCGCTGCTGTGATAAAGCGTGCACCAACAC 120	
QY	121	AGCTCTCTGATGATCTTTAAATGGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCT 180	
DB	121	AGCTCTCTGATGATCTTTAAATGGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCT 180	
QY	181	ACGTGGAGAAATACATATGAATGTTTGAAGAATCCATATGGCCAAACGGGTGG 240	
DB	181	ACGTGGAGAAATACATATGAATGTTTGAAGAATCCATATGGCCAAACGGGTGG 240	
QY	241	TCCGGGCGTTAAACCGGTAAACCGATCACCCTGACGGAAGAAATGCCMAAAGAGTTTCA 300	
DB	241	TCCGGGCGTTAAACCGGTAAACCGATCACCCTGACGGAAGAAATGCCMAAAGAGTTTCA 300	
QY	301	CACCATGGCGCTTATCCACACCGTCTGACATCGAAGCGTGCACCAATTAATGCT 360	
DB	301	CACCATGGCGCTTATCCACACCGTCTGACATCGAAGCGTGCACCAATTAATGCT 360	
QY	361	CGACATCGAGATGCTTTTGAAGGTGCTATCAATTCGGAACAGGATATTCGAGCGAGTT 420	
DB	361	CGACATCGAGATGCTTTTGAAGGTGCTATCAATTCGGAACAGGATATTCGAGCGAGTT 420	
QY	421	GCTAAATATGCGCTTTCTCAACCCACAAACGACCGATCATGCTCATGCTGCGGAGAA 480	
DB	421	GCTAAATATGCGCTTTCTCAACCCACAAACGACCGATCATGCTCATGCTGCGGAGAA 480	
QY	481	AGGATGATGCTCGGTGCTATATCAATTCGATGTTGGCGCGCGCGGTGTTGTCCTGCTG 540	
DB	481	AGGATGATGCTCGGTGCTATATCAATTCGATGTTGGCGCGCGCGGTGTTGTCCTGCTG 540	
QY	541	CATCTCGGCGATGATTCGCGATTTTGGCGAGTGGACCGGAGGACCTGACGCGCATGCT 600	
DB	541	CATCTCGGCGATGATTCGCGATTTTGGCGAGTGGACCGGAGGACCTGACGCGCATGCT 600	
QY	601	CAATGATCGCTGCCAGAAAGAGTGGCGATGATTAACCTCGACAGTGAAGAAGTCTACTG 660	
DB	601	CAATGATCGCTGCCAGAAAGAGTGGCGATGATTAACCTCGACAGTGAAGAAGTCTACTG 660	
QY	661	GAGCAACGCCATACGCTTCCCTATTAACCCCATATTCGCTGCTGCTGCTGCTGCTGCT 720	
DB	661	GAGCAACGCCATACGCTTCCCTATTAACCCCATATTCGCTGCTGCTGCTGCTGCTGCT 720	
QY	721	AATTGACTCAATCAATTCATCGACGCCAGACAATCACGGCGGGAATATGGATGTCGCGGA 780	
DB	721	AATTCACATCAATTCATTCACGCGGAGCAATCACGGCGGGAATATGGATGTCGCGGA 780	
QY	781	TATAGGACACGAGGAGTATACCTATCTCCCGGTAGCTGGCGCTGGAGCCGCTGTTTAT 840	
DB	781	TATAGGACACGAGGAGTATACCTATCTCCCGGTAGCTGGCGCTGGAGCCGCTGTTTAT 840	
QY	841	TGGTGATSCCATGCTGTACAGGGTGATGTPGAGATTTGCGGAGCCGACAGTAGAGTTGC 900	
DB	841	TGGTGATSCCATGCTGTGTAGGGTGATGTPGAGATTTGCGGAGCCGACAGTAGAGTTGC 900	
QY	901	CTCAATCACCACCAACAAAGTCGATTTGATCAAGAACTGGGAGCTTTCCGTGGCCACCA 960	
DB	901	CTCAATCACCACCAACAAAGTCGATTTGATCAAGAACTGGGAGCTTTCCGTGGCCACCA 960	
QY	961	GGAGAAATGCGGAAATATTTAGTATTTGGAGTGGACGTCGCGTGGAGGATGCGAGCGG 1020	
DB	961	GGAGAAATGCGGAAATATTTAGTATTTGGAGTGGACGTCGCGTGGAGGATGCGAGCGG 1020	
QY	1021	AATTGCAATCGCGACCTTAATTTACITGGCTGGTGAAGACCTTTGGCTTGGCAACATGGA 1080	
DB	1021	AATTGCAATCGCGACCTTAATTTACITGGCTGGTGAAGACCTTTGGCTTGGCAACATGGA 1080	
QY	1081	TGCTACATGCTTCTGAGTCAATCGGCAAGTGGGCAACATGGTGGACCCCAA 1140	
DB	1081	TGCTACATGCTTCTGAGTCAATCGGCAAGTGGGCAACATGGTGGACCCCAA 1140	
QY	1141	ATACACCGTTCGCGGATGCTGTAACAAACCTGTTAGTTAGTAGGAATAAATTAACCG 1200	
DB	1141	ATACACCGTTCGCGGATGCTGTAACAAACCTGTTAGTTAGTAGGAATAAATTAACCG 1200	
QY	1201	TGACATATACCGGATGATGATGCGGGTAAATGTTAGTTAGTTAGTTAGTTAGTTAGTT 1260	
DB	1201	TGACATATACCGGATGATGATGCGGGTAAATGTTAGTTAGTTAGTTAGTTAGTTAGTT 1260	
QY	1261	CAGCTAAAGCAGGTGTCATATGGGCGGACATACACCCATCAATATTTGTTTACTTACT 1320	
DB	1261	CAGCTAAAGCAGGTGTCATATGGGCGGACATACACCCATCAATATTTGTTTACTTACT 1320	
QY	1321	TTCAGCGGAGTGGCGGCGGACAGAGTTGTCACATGGCGGAGGACCAACCGAGCTATT 1380	
DB	1321	TTCAGCGGAGTGGCGGCGGACAGAGTTGTCACATGGCGGAGGACCAACCGAGCTATT 1380	
QY	1381	GCGGAAATTAATCAAAATGGCGGATCAACGGCAGACCACTCAATGCAATTCATTTGAT 1440	
DB	1381	GCGGAAATTAATCAAAATGGCGGATCAACGGCAGACCACTCAATGCAATTCATTTGAT 1440	
QY	1441	CC 1442	
DB	1441	CC 1442	
RESULT 3			
AP003583/3 339650 bp DNA linear BCT 28-NOV-2001			
DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 3/19.			
ACCESSION AP003583 BAC000019			
VERSION AP003583.1 G:17129939			
KEYWORDS Nostoc sp. PCC 7120 DNA.			
SOURCE Nostoc sp. PCC 7120			
ORGANISM Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.			
REFERENCE 1			
AUTHORS Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriouchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shampo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M. and Tabata, S. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120			
JOURNAL DNA Res. 8 (5), 205-213 (2001)			
MEDLINE 21595285			
REFERENCE 2 (bases 1 to 339650)			

## AUTHORS

Kaneko, T.

## TITLE

Direct Submission

## JOURNAL

Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, 1-1-1, 1532-3, Kisarazu, Chiba 292-0812, Japan.  
(E-mail: kaneko@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/cyanobase/,  
Tel: 81-438-52-3935 (ex. 2336), Fax: 81-438-52-3934)

## FEATURES

Location/Qualifiers

## Source

1..339650

/organism="Nostoc sp. PCC 7120"

/db\_xref="taxon:103690"

/note="synonym: Anabaena sp. PCC7120"

complement(98..2698)

/gene="all0594"

complement(98..2698)

/gene="all0594"

/note="ORF\_ID:all0594"

/codon\_start=1

/transl\_table=1

/product="aminopeptidase"

/protein\_id="BAB72552.1"

/db\_xref="GI:17129940"

/translation="MPHSYFDDQSNNAHKSFELGAPRHPNPKPQGVHEIFLDLND  
IPQSYEGSGSIRLLPIRNGIDRLTLDAVNOJRSVQVDEVQDDFYDGGVLIQLSQ  
PTEIGKRILIAIYAVKPGQGIYFIQPDGSHPKTVQVMTQGEDESRWFPCFVVP  
GOLSTSEIRVRPKSLMAISNGELIDIOEDGKHKIYHWSQOQIHPTVLM-LAVADFAF  
IRDEWRGKPTVYVVEKGRADAKRSGKTPRIJFELSEKYGVAFAHPKACVGSDFI  
FGGHESTLLIDRCILLDERAALDNKTESLVVHELAWQFGGLLVIKWASHAIKE  
QMASYSEVMWTEQISAEERAYRIILSARSTLSDSGRYRRPMV-HVYREA:FLYDRH  
DYKYSYGHM:RAELGEDLPWAQ-QTFVRDNAHKTVEI:DLLEAI:EKATGRNLAF:LF  
QYVYRGCHPDFKAYSDWGHAKLAKVTVTQVPAKDQKDFDLRIPI:G:GVNQ:EA  
AAQKTQTVVRHEQGFYPLTEKPDF-SFDVGNHMLKTVSLVPIPELAKG:IASGP  
DPIISIVAAALAKGGLLEAKTAA:SLKNEPFWGVVEKALAEIKILQAF:IGYVE  
GLADKNAYVRAYVQALAKTIKILASYAKGVYKVDGASVYESAAALAGA:NAINF  
DLEKAKAIKLIOSLEERAGNVEVRSAGLGLLEAKT:SEAAALDL:LYTKIGVYQ  
PKR:AAIRAGKTI:SPGSSVGNLERIIA:RAGELSKENFFLTIMAYVAALGOMETPKANG  
VLRSLAQTAGRYVRFAEELANVQKNIGTGNALRLKLRFFFCOLKQCNOLKASRIEN  
LEAKSQTKQ"

complement(2862..3059)

/gene="as10595"

complement(2862..3059)

/gene="as10595"

/note="ORF\_ID:as10595"

similar to cytochrome d oxidase chain I"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB72553.1"

/db\_xref="GI:17129942"

/translation="MRHCVHP:IIILSLPLYKIIILN:QWKPSDRPHKVGILYYAFRYK

IAIGFF:LAGLMVSTLFFEV"

complement(3151..3810)

/gene="all0596"

complement(3151..3810)

/gene="all0596"

/note="ORF\_ID:all0596"

unknown protein"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB72554.1"

/db\_xref="GI:17129942"

/translation="MGVCHSKDVSEFC:FGRFMEFV:IKDGYKLGKLLIL:TAGDECYVK  
LSKQLEFDFWRLPTGLLQVYGEKKYVTKLAEAVMAAATETSSAVKQVPAH  
ENATPKPKAKATILVQKQKMGKGLCALEATLSDRGLEQDQVTIKGTGCMKNC  
KAGPNLVMPDKTRYTRICADQVPKLNKHFADSQSDSHSQAVISNTNK:EPFTEVSA  
F"

complement(3877..4134)

/gene="as10597"

complement(3877..4134)

/gene="as10597"

/note="ORF\_ID:as10597"

hypothetical protein"

/codon\_start=1

gene

/transl\_table=11

/protein\_id="BAB72555.1"

/db\_xref="GI:17129943"

/translation="MKQFSVEEEKSVVEVDADRQWQYQRIQLEELPCWCEANQPLK

VELVTPMTVVQLWSVMRQFTSSRQDLIYSLINMQSRYPYS"

4451..5146

/gene="alr0598"

4451..5146

/gene="alr0598"

/note="ORF\_ID:alr0598"

unknown protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAB72556.1"

/db\_xref="GI:17129944"

/translation="MYRTRP:LLSTLS:LVLLQGFV:IVQAGSENTGDVASTAIQOV  
IAANLMTNSADGNFYPERLMNRAELASILVKAFRLNKOBAANOPKSLITVADYPRSHWA  
YODIOTVLTQDVMKGYRGNLFFPNORVTRAEALAI:FAOAYGVFOFSDYVYNEILAPYP  
DANSIPNARRAKAIATVITEGFTINTDAQNNITPLRPMTRGDMAYLISKYLQROQKQPDUT  
PVVPTTIDSPQLP"

5596..7503

/gene="alr0599"

5596..7503

/gene="alr0599"

/note="ORF\_ID:alr0599"

/codon\_start=1

/transl\_table=11

/product="1-deoxy-xylulose 5-phosphate synthase"

/protein\_id="BAB72557.1"

/db\_xref="GI:17129945"

/translation="MHLSEITHPNQLHGLS:IROLQOIAKQIRDKHLQTVAEFGHGLGP  
GLGVELTGLQYOTLD:DRQKVTVDVGHQAYPHKLITGRYSDFHTLRQKDGIAGYLKR  
GENKDFHCAGHASTSI:SAALGNALARDMNGEKFAVAVIGDCAITGGMALPAINHAG  
HLKTNLLVYLNPNEMSI:SPNVGCAIPRYLNKMKLS:PVQOFIKDNFEQFKH:IPFVGES  
LSPLEKIRKEMRLAVPVKGVFEELFTIMGVDGHNLEELIATFQQAHQIAGPVL  
VHVATIKGGLYELAEQDQVGYHAQTFPNLTITGKAIPSNKPKPAYAKVFSHTLVKLA  
CNPK1IGITAAATAGTGLDKLQAKLPQYIDVGAIEQHAVT:LAAGLATGEMRPVAAIY  
STFLQRAYDQI:HDVQICNLPVFC:LDRAIGVSDGPT:HOQMYDIAVLCIPNIVMA  
FKPFAQRMVVTG:IEHTTGP:IAMRFP:RNGYGVPLMEEGWEPLE:IGKGEILRNGDV  
LIITGYSTMYP:SNQAAEILSEHGIEATVINARFKPLDTELIVPLAQKIGRVVY:LEEG  
CYMKGGSVAVEALLDADVVPVVKRIGIPDLVVEHAIPDESKAELGLTSRQIAERVLO  
AYFQKQSPVS"

7793..6281

/gene="alr0600"

7793..6281

/gene="alr0600"

/note="ORF\_ID:alr0600"

unknown protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAB72558.1"

/db\_xref="GI:17129946"

/translation="MADELDKVKIECFILSKLSMKLNFSMKKLSNLGRWIATTVFCL  
SAIAFYWQGAFFADT:SAMADNSTYLIATSDATSLRNKSKFEIRDTQDNVKAARNA  
SRVYDARTNGST:AERKANKDASRIQRAEEDARTETAYVDNNLNAVERTIDNIKGAFT  
NN"

complement(8399..9469)

/gene="all0601"

complement(8399..9469)

/gene="all0601"

/note="ORF\_ID:all0601"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAB72559.1"

/db\_xref="GI:17129947"

/translation="NSIVFRDLLK:YVGSNGHTGESLNRYEAAATATKMMFLGEATPAQI  
GAFLIAHRIKRP:GTEELAGMLDAFELGPKLQPISSKRPVIVFVGIDRTGRTATPLSP  
VTALLMATVEQPVY:IHGGDRLP:TKYGLPLTEIMQGLGVNWTLSLEKTEQVEAOTGIG  
FVYTPKHFP:LTNSINEYRQOLGRPP:LATWELIWCPCYAGDAHVIAQYVHPHTEGMRQV  
ALEBGRVSKFTLVKGLEGSCDLPDRITAIISLSAPELRFVLS:THEYTFGTITKNVPL  
GSTEELTIQDYLQAGKSGELQQTALWNGGFYLWRS:GICADMRSGIAKABELNSGLV  
AAKLQELIQVNSVTENLLSIT"





CY	326 TTGATCAGAA 936	REFERENCE	9 (sites)
DB	438 CTGCACAAGAA 428	AUTHORS	Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T., Hirama,C., Fujii,F. and Takami,H.
LOCUS	AP001507/c	TITLE	Characterization and comparative study of the rrrn operons of alkaliphilic Bacillus halodurans C-125
DEFINITION	Bacillus halodurans genomic DNA, section 1/14.	JOURNAL	Extremophiles 4 (4), 209-214 (2000)
ACCESSION	AP001507 BA000304	MEDLINE	20426005
VERSION	AP001507.1 GI:10172612	PUBMED	10972189
KEYWORDS	Bacillus halodurans DNA.	REFERENCE	10 (sites)
SOURCE	Bacillus halodurans	AUTHORS	Takami,H., Nakasone,K., Takami,Y., Maeno,G., Sasaki,R., Masui,N., Fujii,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and Horikoshi,K.
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillariaceae; Bacillus.	TITLE	Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis
REFERENCE	1 (sites)	JOURNAL	Nucleic Acids Res. 28 (21), 4317-4331 (2000)
AUTHORS	Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fujii,F., Nakamura,Y. and Inoue,A.	MEDLINE	20512582
TITLE	An improved physical and genetic map of the genome of alkaliphilic Bacillus sp. C-125	PUBMED	11058132
JOURNAL	Extremophiles 3 (1), 21-28 (1999)	REFERENCE	11 (bases 1 to 100150)
MEDLINE	99184645	AUTHORS	Takami,H. and Takaki,Y.
PUBMED	10086841	TITLE	Direct Submission
REFERENCE	2 (sites)	JOURNAL	Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group; 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan
AUTHORS	Takami,H. and Horikoshi,K.	FEATURES	(E-mail:takami@jamstec.go.jp, URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html, Tel:81-468-67-3895, Fax:81-468-66-6364)
TITLE	Reidentification of facultatively alkaliphilic Bacillus sp. C-125 to Bacillus halodurans	Source	Location/Qualifiers
JOURNAL	Biosci. Biotechnol. Biochem. 63, 943-945 (1999)		1. 300150
REFERENCE	3 (sites)		/organism="Bacillus halodurans"
AUTHORS	Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y., Masui,N., Fujii,F., Takaki,Y., Inoue,A. and Horikoshi,K.		/db_xref="taxon:86665"
TITLE	Sequencing of three lambda clones from the genome of alkaliphilic Bacillus sp. strain C-125	gene	/note="alkaliphilic"
JOURNAL	Extremophiles 3 (1), 29-34 (1999)	CDS	584..1933
MEDLINE	99184646		/gene="dnaA"
PUBMED	10086842		584..1933
REFERENCE	4 (sites)		/gene="dnaA"
AUTHORS	Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and Horikoshi,K.		/note="BH0001"
TITLE	Sequence analysis of a 32-Kb region including the major ribosomal protein gene clusters from alkaliphilic Bacillus sp. strain C-125		/codon_start=1
JOURNAL	Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)		/transl_table=1
MEDLINE	99209008		/product="initiation of chromosome replication"
PUBMED	10192928		/protein_id="BAB03720.1"
REFERENCE	5 (sites)		/db_xref="GI:10172613"
AUTHORS	Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G., Sasaki,R., Hirama,C., Fujii,F. and Masui,N.		/translation="MENIDHMERALKSMKKVSKPSFETWLTAKTANSLEDSTLIIT APNEFFARDLEKHYDELISITDIDTGVRLYPAFVPISTOLDEFFVEQLKKMKQPP AONGEMNNLNDKYDTFTVIGSNRFAHAASLAAVAPAKAYNPLFYGGVGLGKT ALHMAIGHVYMDHNPNAKVYLSSEKFTNEFINAIRDNKAVNFRNRYRNVDVLLIDDI QFLAGKQTOEEFFHTFNALNDKQIVISDRPPKEIPTLEDRLSKRPFEGWLLIDIT PPDETATILRLKKAENLDIPNEVMVLIANOIDNIRELGALIRVAVYSSLIIND MNADLAATFKDIIIPNAKPRVLTIDTOKTVGEYFHVKLEDFKAKRTKSVAFPROIA MYLSRELTASLPKIGSEFGGRDHTTVIHAHEKISKLLSTDQELQKIQIDSKLRS"
TITLE	Genetic analysis of the chromosome of alkaliphilic Bacillus halodurans C-125	gene	2103..3245
JOURNAL	Extremophiles 3 (3), 227-233 (1999)	CDS	2103..3245
MEDLINE	99411980		/gene="dnaN"
PUBMED	10484179		2103..3245
REFERENCE	6 (sites)		/gene="dnaN"
AUTHORS	Takami,H.		/note="BH0002"
TITLE	Genome analysis of facultatively alkaliphilic Bacillus halodurans C-125		/codon_start=1
JOURNAL	(in) Extremophiles in deep-sea environments (Ed.):		/transl_table=1
MEDLINE	HORIKOSHI, K. TSUGI:		/product="DNA polymerase III (beta subunit)"
PUBMED	249-284; Springer-Verlag (1999)		/protein_id="BAB03721.1"
REFERENCE	7 (sites)		/db_xref="GI:10172614"
AUTHORS	Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.		/translation="MHEVIDRDIFVONNVNHNVSXAVSRRTTIPILTGIKIVADHEGVTL TGSQDISIETFTPLPEGRQNVKQEGSIVLQAKVFAETVKKLPEQETIEIHQDSF VTIIRKSSVFNGLNDPDEYPRLPVLEEDHVERLPQKTLKDLIROTVAFTQETRRP VLTGVNFEIDGILCTATDSHRLAMKRVKPVKNDDLOFSNVVIGKSLNELSKLID ENEELEDIVYTFNQTFLKLNKMLFFSRLLGKYPYTKNMIPKEAKTSFVHTKAFIQI IFRALLSREGKQVNIHLKLGQVVEVTAITPEIGKVTENVAIQEGEELRSTNG KNVIDALKVYVDESIIHIAFTGAMSPFVLSPTDHDQSLHLFSFVRVY"
TITLE	Replication origin region of the chromosome of alkaliphilic Bacillus halodurans C-125	gene	3429..3650
JOURNAL	Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)	CDS	/gene="BH0003"
MEDLINE	99356711		3429..3650
PUBMED	10427704		/gene="BH0003"
REFERENCE	8 (sites)		/note="BH0003"
AUTHORS	Takami,H. and Horikoshi,K.		unknown conserved protein"
TITLE	Analysis of the genome of an alkaliphilic Bacillus strain from an industrial point of view		/codon_start=1
JOURNAL	Extremophiles 4 (2), 59-108 (2000)		
MEDLINE	20263314		
PUBMED	10805564		

```
/transl_table=11
/protein_id="BAB03722.1"
/db_xref="GI:10172515"
/translation="MSQPLPTTSYITIGQLLEKVGVIDTGGMAKNYLAEEHVLNAGE
SEDRGRKLFSGDVLADSVILVDAAE"
3662..4777
/gene="recf"
3662..4777
/gene="recf"
/note="BH0004"
/codon_start=1
/transl_table=11
/product="DNA repair and genetic recombination"
/protein_id="BAB03723.1"
/db_xref="GI:10172616"
/translation="MHIERLLKOFBNYDELVSFEFNVNVIIGENAGGTNVLEATY
FLAKASHRAKELDOWEPAPFIEGAFQKQNGPLSLHVLGSKGKVKVNGLEGR
RISDIIGAVNVMGPELDNLVKGSPQIRRRFDELMEIGMSPVYVLAHMYKILLGR
NHLKQLFGKANSQMLDLVDLIIEVAVEVTKRFEFIQLQWAEIEHQAISRKE
KLVTIYEPVSIVSGLNLSKRGFYQAYEOKKERERQRTTIFGPRFDIIVFVNDK
DVQYSGOQORTALSILKAEIELMKETVGDYPIILLDVLSRLDDYRQSHJHAIQ
HRVOTFVTTNVDGIDHOTLOEATYVEQSGSIKKKG"
4788..5057
/gene="BH0005"
4788..5057
/gene="BH0005"
/note="BH0005"
/unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB03724.1"
/db_xref="GI:10172617"
/translation="MFHILGSDHVIKSLVITLIDPMEQSSGVTKVEYFHHQKNGNT
VTISDLKISIVVPTDIYVSPVSSVTLKRRSHAVSELDRTGESE"
5130..7043
/gene="gyrB"
5130..7043
/gene="gyrB"
/EC_number="5.99.1.3"
/note="BH0006"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BAB03725.1"
/db_xref="GI:10172618"
/translation="MTREQAYDRSQIQLVSLFAVRKRRPGMYTCSSRLGHLVWF
IVDNI:DRAMAGFCDFIGVYKNSITVTONGRQIPVG:HEKMGRAVAVIMVLVA
GKFGGQYKVGSLGIVGASVVALSTMLEVRRECKVHYQKFRVPAADCEVIG
TDRTGTHRPQSDIFTEITVPEYDTLASR:RELAFLNKLGLR:KIDMREFEKSSE
PHYEGGISFVHELNRETTHETPIHIEGEGQGVYVEIAVQYNDGFTSN:YSPANN:
NTHGGTHESGFKTGLTRVINDYAKNLFRESQPNITGDUVRGL:ALISVPTDPG
FEQQ:KTKLGNSEARITLSTFENHREFLIEHPQVARKVDKGMASRAHRAAKAR
ELTRKSALEYSPLGK:ADKSSRDASTSE:YVIEGDSAGSAGKQGRDHFQATPLR
KILNVEKARILKLANNEIRAITALGTGIDQDFIEKARYHKIITMTADVPGAHI
KLLITFYRYMRPLIEHYVIAQPLYKVQGGKIQYAYGDAEQ:ILSQLPDKAK
PGIORYKLGEMNPSQLETTMDPQRTVIQVI:EDAKKA:SEIFELKGLHVFRRPE
IQENARYKNLDI"
7065..9566
/gene="gyrA"
7065..9566
/gene="gyrA"
```

Query Match 4.2% Score 60.4; DB 1; Length 300150;  
Best Local Similarity 44.9%; Pctd. No. 4.2e-06;  
Matches 229; Conservative 0; Mismatches 281; Indels 0; Gaps 0;

```
Qy 639 TCGACAGTGAAGAGTCTACTGGACAAACGACATACGCTCCCTATAACCCCATATG 598
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36944 TCGAAGGGATACGTCGT:TTTCCAGGGTCTCGGCTCTCTTCAACCGA:GATCG 36885

Qy 699 GCACCTTGAGCGTATCGCCGAAGAAATGACATCAATTCATCGACGACACATACG 758
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36884 GAGTGATCGAAGCGGACCGCCAGGAGAACCGATCAATAACGGTACATCCACGACACG 36825
```

```
Qy 759 GCGGGAATATGGA:GTGCGCGGATATAGGACACAGGAG:ATTACCIATCTCGCGGTACGTG 818
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36824 GCGGCAACCTACAGACACGAAAGATATTAAGCCAGGAACAACGGTCTATCTTCAGCTGGAGG 36765

Qy 819 CGCCTGGAGGCGCCTGTTTATTTGGTGGTCCATCGCTTGTGTCAGGCTGATGCTGAGATT 878
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36764 TGGATGGAGCATTA:CTCGCCCTCGGAGATTACATGCCCAATGGAGATGGGAAATCC 36705

Qy 879 GCGGACCCGACATAGAGTTTGGCTCAATCACACCAATCAAGTCGATTTGATCAAGAACT 938
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36704 T:ATCTGTGTGTGTAAT:GTGTCACCGTTTACGTTTAAAGTAATGTCAGAAAGAGC 36645

Qy 939 GGCAGCTTTCCTGGCCACGAATGGAGATGCGGAAATATATGATGATTTGCGCATGCGAC 998
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36644 GAA:GTGTCCTTCTCGGCTTAAACAGATACCCATTTATGACCATAGCTCTGGG 36585

Qy 999 GTCCGCTGGAGATCGAGCGGAATTTGCAATTCGCGCACTTAATTTACTGGTGTGAGAG 1058
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36584 AAACGTAGATGCAGCGCTGTTCAGCGCACAAAAAAGATGGCTACCTTTTAGCGAATC 36525

Qy 1059 ACTTTGGCTTCGACAACTGGGATGCTTACATGCTTCTGAGTCAATGCGSCAAGTGGCGG 1118
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36524 GCACAGCACTGTCGATAGAAAGAGGAGGATGCTTTTGGAGGGTCTGCTGATCTATACG 36465

Qy 1119 TGGGCAACATGTCGACGCCCAATACACCG 1148
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36464 TCAGTCAATCGTCAATCCACTAAAGACAG 36435

RESULT 6
AE006875
LOCUS
DEFINITION
Sulfolobus solfataricus section 234 of 272 of the complete genome.
ACCESSION
AF006875 AE006641
VERSION
AE006875.1 GI:13816157
KEYWORDS
Sulfolobus solfataricus.
SOURCE
Sulfolobus solfataricus
ORGANISM
Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
Sulfolobus.
REFERENCE
1. (bases 1 to 10485)
She,Q., Singh,R.K., Confalonieri,F., Zivanovic,Y., Allard,G.,
Awayez,M.J., Chan-Weiher,C.C., Clausen,I.G., Curtis,B.A., De
Moors,A., Brauso,G., Fletcher,C., Gordon,P.M., Heikamp-De Jong,I.,
Jeffries,A.C., Kozera,C.J., Medina,N., Peng,X., Thi-NGOC,H.P.,
Redder,P., Schenk,M.E., Theriault,C., Tolstrup,N., Charlebois,R.L.,
Doolittle,W.F., Duquet,M., Gaasterland,I., Garrett,R.A.,
Ragan,M.A., Sensen,C.W. and Van Der Oost,J.
The complete genome of the crenarchaeon Sulfolobus solfataricus P2
Proc. Natl. Acad. Sci. U.S.A. 98 (14), 7835-7840 (2001)
21332296
11427726
REFERENCE
2. (bases 1 to 10485)
She,Q., Singh,R.K., Confalonieri,F., Zivanovic,Y., Allard,G.,
Awayez,M.J., Chan-Weiher,C.C.-Y., Clausen,I.G., Curtis,B.A., De
Moors,A., Brauso,G., Fletcher,C., Gordon,P.M.K., Heikamp-de
Jong,I., Jeffries,A.C., Kozera,C.J., Medina,N., Peng,X.,
Thi-NGOC,H.P., Redder,P., Schenk,M.F., Theriault,C., Tolstrup,N.,
Charlebois,R.L., Doolittle,W.F., Duquet,M., Gaasterland,T.,
Garrett,R.A., Ragan,M.A., Sensen,C.W. and Van der Oost,J.
Direct Submission
Submitted (24-APR-2001) Europe/Canada joint project: Copenhagen
University, Denmark; Universite Paris-Sud, France; Wageningen
University, The Netherlands; Institute for Marine Biosciences &
University of Ottawa, Canada
Location/Qualifiers
1. 10485
/organism="Sulfolobus solfataricus"
/db_xref="taxon:2287"
complement(472..828)
/gene="SSO2807"
complement(472..828)
CDs
FEATURES
SOURCE
1. 10485
gene
complement(472..828)
CDs
```





[illegible]

```
/translation--"MKFFELDTANIEBIEKKA:SLGVISGVYTTNFSLVAKGKGFKEVIR
EIAE:VDGPISAEVI:SDHECMYKEARELAKIHKNIVIKIPMTAEGLKAVNLSKEGI
KTNVTITFSANKAZLAARAGATYVSPFVR:DDINTDGMOTFECIVTIL:TNVDIOARI
ITASVRHPIHVI:FAAKLGAHIATVP:KVLHQMIK:PLTDIGIEKFKHKKACLK:"
complement(3193..3637)
/gene="MipB"
/notes="Pfam match to entry Transaldolase, Transaldolase,
score 311.1, E-value 3.40E-92"
complement(3851..3855)
/gene="MipB"
3984..4896
/gene="TTE1919"
3984..3988
/gene="TTE1919"
3997..4896
/gene="TTE1919"
/notes="Best blastp hit - q1110172637 dbj|BAU03744.1|
(AP001507) acetamidase [Bacillus halodurans], score 202,
E-value 3.00E-51"
/codon_start=1
/trans_table=1
/product="predicted acetamidase/formamidase"
/protein_id="AAK25099.1"
/db_xref="GI:20516930"
/translation="MKYSISADYHIFAFSKENKPA:SVKSGDELEVE:TCGTSNQLDS
NEQKIDEMDNHVNPAIGP:FVESAKEGVKIKK:KKEVAEKGVLAIGKGLVLNL
MEGLYSKVVDIKGKVIKFNKALPKPMTGIVGVAPKEGKSGNCGVPSHGGKNDIT:
IAEGAEYFVFVEGALLAIGDHALMGDEGVGVGVAGVAGVLLVEVVGKGNK:KNP
VKTAEATATASAESLDKAVEIAVHDMAELEFKKRTDLSGEGTATLFTSGNNAQISQV
VDP:KTAARFSLPNWILSYGIRF"
4896..5602
/gene="CcmA10"
/notes="CcmA10"
/notes="TTE1920"
4896..4900
/gene="CcmA10"
4910..5692
/gene="CcmA10"
/notes="Best Blastp hit - q113814525|gb:AAK41557.1.
(AE006746) Antibiotic ABC transporter, ATP binding
protein [Salitobus solifarius], score 102, E-value
6.00E-21"
/codon_start=1
/trans_table=1
/product="ABC-type multidrug transport system, ATPase
component"
/protein_id="AAK25100.1"
/db_xref="GI:20516930"
/translation="MTAEIIVKLENTVTKLGRRLTLEDINFEVYKGEKVKILGPKKSG
KIALVEITLGLXPDGASVSLGTDPTNSSL:GOASVYKNSFPLHWRVQVINF
KKOLWGENSGHIFALDISNMKNTTVGKLESEERAKLNLFFALTAKALITLGDPA
NSLFDLEIKNPFHTSITTSKTPPVYDIADRIYVILHEGRIKKSIAKQDYSSLEK
LKS:LYOYAPED"
5003..5515
/gene="CcmA10"
/notes="Pfam match to entry ABC_tran, ABC transporter,
score 21.5, E-value 3.53E-07"
5586..6287
/gene="TTE1921"
5586..6287
/gene="TTE1921"
/trans_table=1
/product="hypothetical protein"
/protein_id="AAK25101.1"
/db_xref="GI:20516931"
/translation="MRUKTYEGVYEVKIRADALKVVFSTLSSIEIRFSSIFQVPL:
LLGFGYAGKGRKLYPDFET:LWMLSSGMAJYIATIPWGAALRERQSNIR:MLSG
VSPHEW:GKSFALLSGISYIAILISPLGYKLNPTWGLVFWPAPFLGFFIILIS
LTLFLCAVGLSVAGTSMAFTISYTLFFGLLFGYISPKRSSEILILFLFLFLF:KSF
LVAKRFEKEMAFY"
5291..6316
/notes="putative rho-independent transcription terminator"
complement(6324..7776)
```

```
/gene="PotB3"
/notes="TTE1922"
complement(6324..7763)
/gene="PotB3"
/notes="Best Blastp hit - q116759463|emb|CAB69791.1|
(AJ237785) putative amino acid transporter [Bacillus]
Query Match 4.1%; Score 58.8; DB 1; Length 7983;
Best local similarity 46.9%; Pred. No. 6.6e-06;
Matches 183; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
QY 646 TCAAAAGTCTACTAGCAGCAACGACATACCTCCCTTAIAAACCCTATTTGSCACCTT 705
DB 434 TGGAAAGGTATATTTAATGAAAACTASCTCTCTCTTAAGCAATGATAGAGTAAT 4400
QY 706 GAGCGATCGCCAGAAATTTACTCAATCAATCTACTGACGCGACACATCAACGGCGGAA 765
DB 4401 AGGAGTTGCCCAAGGAGGAGGAGGATTAACTGCGGACGCGAGGTCCTCCACGTTG 4460
QY 766 TATGATGTGCGGATATAGGACACGAGGAGTATTACCTATCTGCGGTTACGTGCGCTTG 825
DB 4461 CATGACACACACTTAAATAGCAGAGGAGGAGTCTACTTCTCTGATTATTAGAGG 4520
QY 826 ASGCGCGCTGTTTATTTGTTGATGCCCATCTGTCAGGSGTGATGCTGAGATTTCGGGAC 585
DB 4521 ACCTTTGCTGCTAGGGGACTTACACGCTTTAATGAGAGATGAGAGGATAGCGCTTTC 4580
QY 886 CCGACTAGAGTTTGCTCAATCACCACCAATCAAAAGTCGATTTGATCAAGAAGTGGCAGCT 945
DB 4581 AGCGCTGAAGTGTGCTGGAAAGTCTTTTAGAAGTGAAGTCAATAAGGATTTGAATCT 4640
QY 946 TTCTTGGCCAGAGTGGAGAAATGCGGAAATATTATGAGTATTGCGACIGACGTCGCT 1005
DB 4641 CAAAATCTCTTGTGAAAGTCTGGAAGTAACTAGCTACTATAGCTTCAGCGCAATCTTT 4700
QY 1006 GGAGGATGCGACGGAATTCGATATCGCGA 1035
DB 4701 AGCAAGCAGTGGAAATAGCGGTACACGA 4730
RESULT 8
LMFLCHR36_01/c
WPCOMMENT
Sequence split into 36 fragments LOCUS LMFLCHR36 Accession AL499624
Fragment Name Begin End
LMFLCHR36_01 1 110000
LMFLCHR36_02 100001 210000
LMFLCHR36_03 200001 310000
LMFLCHR36_04 300001 410000
LMFLCHR36_05 400001 510000
LMFLCHR36_06 500001 610000
LMFLCHR36_07 600001 710000
LMFLCHR36_08 700001 810000
LMFLCHR36_09 800001 910000
LMFLCHR36_10 900001 1010000
LMFLCHR36_11 1000001 1110000
LMFLCHR36_12 1100001 1210000
LMFLCHR36_13 1200001 1310000
LMFLCHR36_14 1300001 1410000
LMFLCHR36_15 1400001 1510000
LMFLCHR36_16 1500001 1610000
LMFLCHR36_17 1600001 1710000
LMFLCHR36_18 1700001 1810000
LMFLCHR36_19 1800001 1910000
LMFLCHR36_20 1900001 2010000
LMFLCHR36_21 2000001 2110000
LMFLCHR36_22 2100001 2210000
LMFLCHR36_23 2200001 2310000
LMFLCHR36_24 2300001 2410000
LMFLCHR36_25 2400001 2510000
LMFLCHR36_26 2500001 2610000
LMFLCHR36_27 2600001 2710000
LMFLCHR36_28 2700001 2810000
```

Genbank release 109; EMBL release 56.0; SwissProt release 38.0; PIR-Protein release 62.0; and OMC release 31.4.  
E-mail address for comments and questions: kyutaka@nie.go.jp  
ORF organization, sequence alignment and more information are available at W.W. Site of Biotechnology Center.  
URL: [http://www.bio.nite.go.jp/donan/genome\\_list-e.html/](http://www.bio.nite.go.jp/donan/genome_list-e.html/).

```
Location/Qualifiers
1. .263050
/organism="Sulfolobus tokodaii"
/strain="7"
/db_xref="taxon:111955"
complement(526..1632)
/gene="ST0001"
complement(526..1632)
/gene="ST0001"
/codon_start=1
/transl_table=11
/product="368aa long hypothetical protein"
/protein_id="BAB64942.1"
/db_xref="GI:15620946"
/translation="MSVIVYTVIIGEINVKDQKFFSNLATAVNIISOSFDEPLEKX
LHEYKDVTEGDEVDKSDSPFYASILSHKVFEDDFNNGRLKALDEEVEKDIERKIK
ELIKRGEADKDEPLWVPLWVSKFDPDILSVFKLDSEFEYIGIKVLTPIKPYIS
ELIYGADPLFFTKNKLPGKVFQSDVDPYIRVQLDNEERENFVLTLEKILGLKT
ELQIRLRINRSEPLNTEYLLIALRYAHWILRTSKLTIGEAVNCPLILLASVDKPTQ
FVKETKDIYHRIENGSGVDSLEIRKEIENLGNWYNTILPSRANTQSVREVKNKINELFKP
MKLGNPTIMVILCHSILIVYKVGVDPKVFRV"
complement(1883..2173)
/gene="ST0001"
complement(1883..2173)
/gene="ST0001"
/note="PIR:C71147 percent identity: 36.264 in 91aa."
/codon_start=1
/transl_table=11
/product="36aa long conserved hypothetical protein"
/protein_id="BAB64943.1"
/db_xref="GI:15620947"
/translation="MEVSNVNMKKVLTRELTEODAIAKAVTIIEKYLPLLILNKSVN
LIRAME:SVKEIKIPYDSLY-ALAEKRGSKLVTGDKKQHEIAKKYVISEI"
2263..2496
/gene="ST0002"
2263..2496
/gene="ST0002"
/note="OML:A00972 percent identity: 28.070 in 57aa."
/codon_start=1
/transl_table=11
/product="77aa long conserved hypothetical protein"
/protein_id="BAB64944.1"
/db_xref="GI:15620948"
/translation="MRQLSSLLNSALLELVGLVWVLSLFFKISFFNSILL
IKFSISSFHLIYFSTISRNLSILKLYTFIL"
complement(2268..2537)
/gene="ST0003"
complement(2268..2537)
/gene="ST0003"
/note="PIR:H69325 percent identity: 40.580 in 69aa."
/codon_start=1
/transl_table=11
/product="89aa long conserved hypothetical protein"
/protein_id="BAB64945.1"
/db_xref="GI:15620949"
/translation="MSKTEFTFSPVYVYVMKTVYSIRIDKELREEMEKYNIKWNEEI
ENFPRREEUKKEIKLKNKINELQTMPTPEINSSSAFLVHEDRNN"
complement(2722..2937)
/gene="ST0004"
complement(2722..2937)
/gene="ST0004"
/note="PIR:S74019 percent identity: 84.058 in 69aa."
/codon_start=1
/transl_table=11
/product="71aa long conserved hypothetical protein"
/protein_id="BAB64946.1"
/db_xref="GI:15620950"
```

```
/translation="MGKITGGSDIDVLLVYKRNEDRNKAJLDFFEYVERELGEKVSYL
FDVKVIYEDERELPPFKSFLRDVRIK"
3144..3401
/gene="ST0005"
CDS
3144..3401
/gene="ST0005"
/Note="Swiss_Prot:P37413 percent identity: 52.00 in 25aa.
motif: phenyl group binding site (MAX box)"
/codon_start=1
/transl_table=11
/product="85aa long hypothetical protein"
/protein_id="BAB64947.1"
/db_xref="GI:15620951"
/translation="MSFSLVILRKVAYPFLSKVYNKESNLIPWVQSLFSLKSLGK
VWLKHYLTLPHKGLAEIYVLKILSSPFLSKSLQQFCN"
complement(3247..3399)
/gene="ST0005"
CDS
complement(3247..3399)
/gene="ST0005"
/Note="P1R.S74018 percent identity: 58.696 in 46aa."
/codon_start=1
/transl_table=11
/product="50aa long conserved hypothetical protein"
/protein_id="BAB64948.1"
/db_xref="GI:15620952"
/translation="MSYNAIEFLPRADYLKASLILFQQGHIYASALDSEVSARLF;
NFLNLS"
3744..5483
/gene="ST0002"
CDS
3744..5483
/gene="ST0002"
/Note="Swiss_Prot:P55577 percent identity: 37.004 in
227aa."
/codon_start=1
/transl_table=11
/product="579aa long hypothetical prolyl endopeptidase"
/protein_id="BAB64949.1"
/db_xref="GI:15620953"
/translation="MDEYEYLN;SDPTISFIEKENKETSKLKKKAMELYPLLELM
DKPEYVSMFAYDEENPAILLYGEKSQLLQNKTIYIPPEGVASEINKVYNKSEIGV
STYERKSDKIITLLSPGKIRLEGWVSEPFYKGECLYIKSVYSPDPGGYDPA
RVEFGDELYVGDMPKPGFVTIKVFDDFLTLVRQKGRYGLYVGEGLDRLKVDGE
VIDIDFOGVEYVOKNAVALNGNKKVEVDYPLVGVSHIGDKTAVEVKEYRPLIF
VDLKKKIGEEVHDNITPMDCKGHTLFLVPSFNKFRVVKXSDGKREVLMOGNYD
VTVKOLYGVGLVHGFVLSRANPKGVTVYGGFRIPLLPSLFSVNRVLINSEYI
LITNIRGGYENGEEHMKAGMLNKKNVKDFAEFLRLVKMGCKAVAMGGSGGLVG
ATINPEYELIDCAVLGHPLVDMLRYDKLYGVKYVVEYGDNDPKYTFYLLSYSEYH
LKKGLPTFYVTGINDRVIHPAHALKYVAKSKSLGNDVNLVFNLSGSIADPESKARE
ESYVVSFIIECLK"
complement(5570..5819)
/gene="ST0003"
CDS
complement(5970..5819)
/gene="ST0003"
/Note="Swiss_Prot:Q40469 percent identity: 32.816 in 67aa.
motif: ATP/GTP-binding site motif: A (P-loop)"
/codon_start=1
/transl_table=11
/product="949aa long hypothetical protein"
/protein_id="BAB64950.1"
/db_xref="GI:15620954"
/translation="MICSSST;SSLDLVFKFSESQVPTTTNNVKFPLVLTLC;SPN
VRLCDLNNYOHGLFALNKIMGKNLVVNAFTGGKTEIWSVYLEMLKGFENLAIY
PTKALAGDGLERIVKYYIENGESVSKOVKTKIVEYGVGVKIKGVSYN;IQSVKR
AKTLTNPEVKNALFONHKHISDFLKKV;IYVDFEFGSKSVTLITLTKIDKP
GIKPOIVMSATLSDOPEVVKPFDFVEIIGKSEFANDYIVLGRDELNLSKALNY
SFDD;VKNFEKYASDKELIYQLFKDQAOLFLYLKLGKDELTIFSRISFAN
NLVGLKGVNTYDGTAPVAHVHSGIDKYVROOVENDMPSGLKVVVTKYKGGIDVS
NVTYVHVGIPIPSVEFGKRGKRMASIKTESVIFPLSLSDVALLDEDSTLKEV
LSLGSFSLILLPNFEKLIDV;RGILNDREFLKSGLTCTIPIQISFELKHKVPL
LFCGNSNVOKNLNDVVERKQGLCIDPMLNALVVRNSVNGKRHVLEISDFSSLM
CNHNLSSVKNALYAYEDICFSWKQHPDMLNDIERGKVMKSLDVLFEEDGFKFQ
VHEIARKFPWYIFSRKVGKIOEYKIELEIYTPSPNIKYDFTLYVYASELEPDDVYK
DKGNVFI;LALLRLNYGIDGLINFGVNSGILKINWESEPTALLKELRMKKVKYRGTE
```

```
LDCKLLDDVRNQPFRRLVLVLYLDFYTFNNDFNEVKAIARFVHYLCNTPIQL
QVISTMPKSPILRSIIIDSFSGKYAVSSPTGEISIFDNEEERAVKAAKLNIETDA
IVVPIYGSICKYKQNEVYIDKEINRLGGPIIPSKFRELVLNCDLSLKEENADN
SITRGEVEHITKIPKRAETIRLMINLWKAYLTDK"
9382..9549
/gene="ST0007"
CDS
9382..9549
/gene="ST0007"
/Note="OWL:AF072619 percent identity: 27.500 in 40aa."
/codon_start=1
/transl_table=11
/product="55aa long hypothetical cytochrome b"
/protein_id="BAB64951.1"
/db_xref="GI:15620955"
/translation="MKACTPSLKHENFIELIVMI;TPISLKRSSRLDIT;SPSRORQSI.
CILNIYELFYI"
complement(9819..10127)
/gene="ST0004"
Query Match: 3.94; Score 55.8; DB 1; Length 263050;
Best Local Similarity 52.94; Pred. No. 0.0001;
Matches 120; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
Oy 717 STATCGGCAGAAA;TGA;TCAATCAATTCAC;TGACGCGCAGACAAATCACGGGGGAATAIG 769
25 172729 SCNTTACNTATAGAGAAATTTAGTACCACTCTCTAGAGAAACGGAGGAATAIG 172670
Qy 770 GATGCGCGGATATAGGACGAGGAGTATTACCTATCTGCGCGGACGICGCCCTGGAGGC 829
2b 172669 GATATTAAGCACITTAACCTATAGGCACATAATATTACCTTCCAGTATTGTAAGGGAGCA 172610
Qy 830 CGCCTGTTATTGGTATGCCCATGCTGTGTCAGGGTATGTTGAGATTTGCGGACCGCA 889
Eb 172609 TTGCTTCCATAGGTGACACTCACTTACCCCAAGTGATGCTGAAGTTTGGGAACGTCA 172550
Qy 890 GTAGATTGCTCAATCACCACCAATCAATCGATGATTTGATCAAGAA 936
Eb 172549 ATAGAGCCCCCTGGAGGTTACANTGAGGTTAGGTTAATAAAAA 172503
RESULT 10
SME591789 294800 bp DNA linear BCT 05-JUL-2002
LOCUS Sinorhizobium meliloti 1021 complete chromosome; segment 8/12.
DEFINITION AL591789 AL591688
ACCESSION AL591789.1 GI:15074950
VERSION
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 294800)
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Gedrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetecole,D., Puchler,A., Purnelle,B., Ransperger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
ANALYSIS of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
JOURNAL 21395507
MEDLINE 11481430
REFERENCE 2 (bases 1 to 294800)
AUTHORS Gouzy,J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EJ Consortium
MELILO EJ Consortium:
Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR2115-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
```

Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,  
D-33615 Bielefeld, Germany. Unite de Biochimie physiologique.  
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,  
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte  
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,  
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr  
<http://sequence.toulouse.inra.fr/mellor1.htm>.

## FEATURES

```

source
    location/Qualifiers
    1..294900
    /organism="Sinorhizobium meliloti"
    /strain="1021"
    /db_xref="taxon:382"
    complement(147..1382)
    /gene="SMC04259"
    complement(147..1382)
    /gene="SMC04259"
    /function="cell processes; transport of small molecules"
    /note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE PERIPLASMIC BINDING ABC TRANSPORTER
SIGNAL PEPTIDE PROTEIN"
/protein_id="CAC46508.1"
/db_xref="GI:15074951"
/db_xref="SPTREMBL:Q92P64"
/translation="MNLRFALAAAGCATTAALPFGAASATDLEVIHWMTSGSEAAVAFI
AKAFDATGNKWDGALAGSGGTARFIM:SRITIGDPMATOFNHRGROAEELVOAGLMR
DLTDTATKENKEIVKPSLLDSCTIEGK:YCAPNLI:SNQMWLSNNAFKGACVEYP
KNWDEVAARALEKAGSIVPLAVGGOPHOANGAFVLMVATAGAKENFEKVFPAQDEEV
RAGPTIAKVFRAADGAKSAG:NVDMNQNATNVTIKAGGOGINGWAGGFEPLA:Q
KAGVDTGLGVLGVNEVISTGDAPFPP:LEDEEKSKAQEVLASIL:KPE:QVAPNLK
KAGSVRGVDI:AAANDCKMKGIDILAKGNVIQGT:DQLLSADSOKOKEDLSEPPFANI
SMTPEDAKRRFDI1AAAD"
    1858..2928
    /gene="SMC04260"
    /feature="miscellaneous; not classified regulator"
    /function="miscellaneous; not classified regulator"
    /note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAC46509.1"
/db_xref="GI:15074952"
/db_xref="SPTREMBL:Q92P63"
/translation="MTGMDGKTKRATTAAPVEGRPTLKTIAFTMGLGVTVTSKALK
DAPDAGTKERVLAKVQIQPNRAGVRLRTGKTVNIVSIW:LEEEIMGITSPLVLI
GTTELACTGVHLVVTYSSAKDPLGPIRYILDTGAADGVILISKTEPNDPNTIL:TER
RPFATHGRTENGLHPYDFDNDEFAYEAVRKLVDGRRLRLLEPPNLT:HSIMR
TGFERGVDFGNANV:FHQVNDISIVAI:RDAPFKLRSPDAPDGVSSGSGAJALI
AGIEAAGKKVGDGEMVSKVPSDFLRLWRPEV:MYEDIRIAGRELAKAVTGHTEGRP
PTTOSLSOPEQPPMARSAVS"
    2954..3023
    /note="Sm-4 OR SMC04506
REPEAT SM-4
Predicted by Homology"
/evidence=not_experimental
    complement(3036..3302)
    /gene="SMC04261"
    complement(3036..3302)
    /gene="SMC04261"
    /function="miscellaneous; unknown"
    /note="Product confidence : hypothetical

```

```

Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL PROTEIN"
/protein_id="CAC46510.1"
/db_xref="GI:15074953"
/db_xref="SPTREMBL:Q92P62"
/translation="MKFTVIFGGKAAEVTAGSVTIRDEARRIKAAAGVEQWRVRSIA
TGAPIPGPIQYLKMOIDFVAGKLEQLNPISDFTDKYPWPAAGI"
    complement(3407..4837)
    /gene="gnd OR SMC04262"
    complement(3407..4837)
    /gene="gnd OR SMC04262"
    /EC_number="1.1.1.44"
    /function="small molecule metabolism; energy metabolism,
carbon: oxidative branch, pentose pathway"
    /note="Product confidence : probable
Gene name confidence : putative
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE 6-PHOSPHOGLUCONATE DEHYDROGENASE
(DFCARBOXYLATING) PROTEIN"
/protein_id="CAC46511.1"
/db_xref="GI:15074954"
/db_xref="SPTREMBL:Q92P61"
/translation="HQAEGTIGLIGLVMSGNLALNAEKGNITAVFNPTVDATRKFYA
EGALQKQIVPCTIEEFVAARPPRPIIIMIKAGDPVQOQNEALKPLKADIMIDA
GNANFDIMRRDALKDGLTIFIGCVSGSGAGHPGSPIMVGGTEESTRRVEKVLTS
IAKYDSDCVAMLVGENGAGHFVKTIHNGIEYADMOMIAEYIGLUDGLKMTAQEIGE
VFGANKRGLNSYLIETEKVKAADPLT:GKPIVDMILDKAGCKGTGKMSVIEAONMG
VPATAEAAVAARSISSAKEEREAKEVLGPPVGEIKVADRGTGKIDENALLAAKI
GAYAQGFAYNAARSKFEFNLMPTITAKIWRAGCIIRSFDELDTITATFKAPDAANLI
VTPFAAMVKESDGLARRVYVSTAVLGELPVPALASALGYFDSYRRGRGTANVIOQRD
FTGAHGFDVRDGDADSHHGPWGSGLNA"
    5273..6691
    /gene="SMC04263"
    5273..6691
    /gene="SMC04263"
    /function="cell processes; transport of small molecules;
amino acids, amines, peptides"
    /note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE AMINO ACID CARRIER TRANSMEMBRANE
PROTEIN"
/protein_id="CAC46512.1"
/db_xref="GI:15074955"
/db_xref="SPTREMBL:Q92P60"
/translation="MDTIIGLNTIFWGVVLYGLLAVGLYFTVRLGFPQIVHFSEMF
RV:SRGSKDAAGISPFQALMVSLSRGTGNLAVVALYLGPGATFMMWIVAFVGA
MATAYAESALAOIYKIRNEDGOYRGGPAPFVIA:RGKNAPAAATIFSACILISFGI:VFN
VOANSTADAVOGAFVGPVLAVGVAVALLSGVIFGGIROIARVAETIVVPMAAYLI:IT
AVYVLANASAVPGVLTWLT:ISSAFGEQAEAGGIIGGIAAMLVNKRGLFSNEAGMGS
APN:RAAVATVPVHPHSSQGFVQSLGFTIDILICTATSVMLLSFLPESGVTGTOL
TOAMSTHIGSAGTIFIAIAIFFAFTISIGNYSYAENALT:YLGGNRGLGIMRCAT
LANVVMGAYESITIVFDAADASMGMLATINLVAILLLSGTIAKLTQDYFEQRKAGAVP
VFHAADYPELOKRGIDGINSRD"
    6773..7663
    /gene="SMC04264"
    6773..7663
    /gene="SMC04264"

```







Road, Exeter: EX4 4QG, United Kingdom

On Dec 16, 1999 this sequence version replaced gi:5224591.

Notes:

Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.

(URL: <http://www.sanger.ac.uk/projects/S.pombe/>)

During 1995 to 1996 about 66% of *S. pombe* chromosome 1 was sequenced by the Sanger Centre. The sequencing of the *S. pombe* genome is now being continued with funding from the European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.

Protein coding regions (CDS) have been predicted with the help of computer analysis using the GeneFinder program in PomBase (an ACEDB database), with additional predictions for the branch-acceptor sites supplied by the program SplicePic. CAUTION: It is possible that for any individual CDS we may have under-estimated or over-estimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.

CDS are numbered using the following system eq SP6C25H2.CDS.SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), C (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Location/Qualifiers

Source

1. .35346

/organism="Schizosaccharomyces pombe"

/strain="972h"

/db\_xref="taxon:4896"

/chromosome="I"

/map="IR"

/clone="cosmid c869"

/complement(1..1610)

/gene="SPAC869.11"

/complement(1..1610)

/partial

/genes="SPAC869.11"

/note="SPAC869.11, len:536, SIMILARITY:Schizosaccharomyces pombe, ISP5\_SCHPO, sexual differentiation process putative amino-acid permease isp5., (543 aa), fasta scores: opt: 1812, E():0, (53.9% identity in 492 aa)"

/codon\_start=1

/label="SPAC869.11"

/product="amino-acid permease"

/protein\_id="CAB60021.1"

/db\_xref="GI:6224602"

/translation="MEPEVSVTSDDTKSLNESKSSLDKVPKSLSEKSYITPGLVYD VEPKGNFVRFDDPKPAKATGEDSTALKRSLKSHMQMISIGAGTGLYVGSGSS LADGPASVIVINSLIGIMHFTIVVAGEMAVYPVAGGNTYATRFIDFANGFAVSW NYFINVFPELELTTCATTFRYWTDINSAAN-STPLVVIIVNI.FGVAYAGVVFIL STVVCVATGFI.LAIILNCGVPIDHRYIGGSIKKHPRGPGFPCSTYIAAFS FSKTEVIGLAADVMPKALPHAVKQVPRIAIEVVSLLIGLISPDPLNMGNG STSVSPVLAIEANIKGLPSFVNAV.IISVSVVINSSTYTAGTGLHGMANKQKQSP FKTDRGLRPLAMIVLLFGFAYINEADKNGVSDVFNKLALSGESNFTVGS ICILHIFRLAKQKGSKELGFSVPMGIWGSICGLGFNILLCKAQFYVSLFPGGX PNANDFGYLAAPVTLAFVIGYKIDRS"

1. .1141

/note="Overlap with: c922 S. pombe chromosome 1"

/complement(3..1394)

/gene="SPAC869.11"

/note="Match to PF00324 aa\_permeases, Amino acid permease score 820.26"

/genes="SPAC869.10c"

/note="SPAC869.10c, len:552, SIMILARITY:Emicella nidulans, PUTX\_EMENT, proline-specific permease, (550 aa), fasta scores: opt: 1968, E():0, (53.3% identity in 520

aa)"

/codon\_start=1

/label="SPAC869.10c"

/product="amino-acid permease"

/protein\_id="CAB60020.1"

/db\_xref="GI:6224501"

/translation="MDEKALQYEEESKRAREJTDIELVSGGIDVEKKYGETKRAIKS RHVQLIAGGCGICTGLFVSGSALSSESPASFLSYVIMSVIIVNMGEMCTYLP LSGASPTIYIBRYVDASIAFAAGNNYAYVFLVASEVTAASIVIEWTYAYPPAGKI AILFLVAVLNSFEVKGFEFTWFAIKVIAIVGLIILGVVIFFGTTPHDLGFRY WKHLAFREYIVKGSRGVGFSAVSKGPAF.LAPELV.FSGETVFAPRNIPKAI STVIRLI:FFYIFGSLTIGVITSSKDPRLNNAISSGASGAASPEVIGIQAIE.PVJN HINAVILTSACSSNSHLPAGSISYSLAKEHOAPKIFKYCNRWGVVIVSVATVLF ACLAFARASAAVFWFCNLSTLSGLAMICVLYAYLOPKAMILNNLWETRPYKT PMSFYATYLLLELLALITLNGFVFGVGTAGNTAAIYITLIFLVLYVAHKLWR NWSFGKRIEEIDVTIGVAEAELEOMYPAPVPRNIIEKFWIA"

3816. .5228

/gene="SPAC869.10c"

/note="Match to PF00324 aa\_permeases, Amino acid permease score 118.22"

/complement(6788..7138)

/note="Tf1 ITR"

/complement(7524..7874)

/gene="SPAC869.09"

/complement(7524..7874)

/gene="SPAC869.09"

/note="SPAC869.09, len:116, SIMILARITY:Schizosaccharomyces pombe, YAAH\_SCHPO, hypothetical 15.4 kd protein c24g7.11c in chromosome I., (143 aa), fasta scores: opt: 192, E():2.5e-06, (39.6% identity in 139 aa)"

/codon\_start=1

/label="SPAC869.09"

/product="hypothetical protein"

/protein\_id="CAB60019.1"

/db\_xref="GI:6224600"

/translation="WPDZAHIIAGHKAALSHPHYSEAKRRARKYLKEHGESHSYTTG TTRGOKADADDAGELRFRSGCTKQFEDNESCAKNIGVRCGYKAAHNPVKSQKGRR HAKSLREVDDESK"

/complement(8736..9448)

/gene="SPAC869.08"

/complement(8736..9448)

/gene="SPAC869.08"

/note="SPAC869.08, len:230, SIMILARITY:Triticum aestivum, P1MT\_WHEAT, protein-l-isoaspartate O-methyltransferase, (230 aa), fasta scores: opt: 602, E():2.2e-32, (48.5% identity in 200 aa)"

/codon\_start=1

/label="SPAC869.08"

/product="putative protein-l-isoaspartate O-methyltransferase"

/protein\_id="CAB60018.1"

/db\_xref="GI:6224599"

/translation="NWSFNLSNAALVOHLVESKFLNORAIKAMNATRSFYCPLS PYMDSQSIGVGTISAPHHAIALORLEPVLQPCSLDIDGSGGYLVAAARVAP NCTVGEIIPOLVETSKKLLKDNHDELMEMYKERRLOINVDGDMGTSEDEKFD ATHVCASASE.PQKLVDDOLKSPGKILIPIGTYSQNIYLIEKNEQCKISKRTLFPVRYV FLIDSPDSSSDY"

/complement(8783..9427)

/gene="SPAC869.08"

/note="Match to PF01135 PCMT, protein-l-isoaspartate(D-aspartate) O-methyltransferase (PCMT) score 263.14"

/complement(10265..11575)

/gene="SPAC869.07c"

/note="SPAC869.07c, len:436, SIMILARITY:Mortierella vinacea., Q02402, alpha-galactosidase, (417 aa), fasta scores: opt: 1443, E():0, (52.3% identity in 407 aa)"

/codon\_start=1

/label="SPAC869.07c"

/product="putative alpha-galactosidase"

/protein\_id="CAB60017.1"





```

JOURNAL Nature 404 (6777): 502-506 (2000)
MEDLINE 20222556
PUBMED 10761919
REFERENCE 2 (bases 1 to 349061)
AUTHORS Parkhill,J.
JOURNAL Direct Submission
TITLE Submitted (30-MAR-2000) Submitted on behalf of the Wellcome
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
URL: http://www.sanger.ac.uk/Projects/N\_meningitidis/.
FEATURES
source
1. 349061
/organism="Neisseria meningitidis 22491"
/strain="22491"
/db_xref="taxon:122587"
/note="serogroup: A"
complement(24..206)
/note="ATR repeat: hmufs hit to HMM ATR (1 - 183). score:
310.39"
/label=ATR
209..212
220..681
/gene="NMA0368"
220..681
/gene="NMA0368"
/note="NMA0368, probable integral membrane protein. len:
133 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=1
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/db_xref="SPTREMBL:O9JWH7"
/translation="MQEQNRKSSFP:VMLLYSVALKIASLSNVAFYLNHGSMEGLIV
LILGSFASLDIRYAGVANYVWLAITVLEALRKAVVPVHAFWGLA:VAFSVKVVY
DRAQNTSDIVRGYAGFYLMWYAFVAVASITGAGNKKRKAASAADGTRNDV"
671..1492
/gene="hemK"
671..1492
/gene="hemK"
/note="hemK"
len: 273 aa; similar
to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa). fasta
scores: E(): 0, 42.3% identity in 279 aa overlap. Contains
P500092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=1
/product="hemK protein"
/protein_id="CAB83570.1"
/db_xref="GI:7379122"
/db_xref="SPTREMBL:O9JWH6"
/translation="HTFDKWLGLSKLPKNEARPHLIQYVSEKTRVQLITRGGERMTDFV
RQADRLAQRIINGSPVAVIIGAREFYGRRTFVNSVLIIPETEH:WEAVCARIPEN
GRVMDLGTSGAVAVTARDPDAVRAISPPALERTAKNAADLGARVFAFGYSWF
DTPMSEGGKWD:IVSNPPY:ENCDKHLISQGLRFEPCALIDFSDLSCLTLAGAP
DRLAEGFLLH:HGPDGAAVAVGVLAENGFSQVET:PLDASLDRTVLSKYNKHLK"
831..840
/gene="hemK"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1162..1171
/gene="hemK"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1187..1207
/gene="hemK"
/note="P500092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
RBS
1565..1568
gene
1578..2966
/feature="NMA0370"
1578..2966
/feature="NMA0370"
/note="NMA0370, probable integral membrane protein. len:
462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
P44640 hypothetical protein H0325 (450 aa), fasta scores:
E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=1
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/db_xref="SPTREMBL:O9JWH5"
/translation="KNVAVVAVIMLVLSLSRVHVVLSITIGAFVGGAVAGMPLONIA
DRAQVSQAGIIPVFNKNGLEGGAKTALSYMLGAFAMATHSGLPQQLAGAVVRKLNK
GMPDSVRSGEGVKKLLSIIILMGMMSONIPIPIAFIPVPLPPLLVFNRLKIDR
RLIACVITFTLVNTYMLPYGFAIFLNEILLGNHSAPOIDVKNINVMAMAFIQL
GLIAGLLAFVHYKPKLYQSNNAIDTGNADNANRPPQSAFSLAAVAIAVCFAPLQ
MVEDSLVGLMGFAVFMVGLVNRKANDVFEGIKMMAMVGFITLAAQGFVAVNNA
TGHIOPLVSSMAIFGNSMAALMLVYGLLVTMGIGSSRSTFLIIRAIYVPLCVGL
GESPLATVAVCTAGALGAGSPASDSTLGTPTMGLNADQHDHRSVIPFIHINIP
LLIAGWIRAWVL"
2967..3191
/gene="slxX"
2967..3191
/gene="slxX"
/note="slxX"
/note="NMA0371, slxX, SLX protein homolog, len: 74 aa;
similar to SLX_HAEIN P44759 SLX protein homolog (73 aa),
fasta scores: E(): 0.0017, 33.8% identity in 74 aa
overlap, and SLX_ECOLI P30857 SLX protein (72 aa), fasta
scores: E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=1
/product="SLX protein homolog"
/protein_id="CAB83672.1"
/db_xref="GI:7379124"
/db_xref="SPTREMBL:O9JWH4"
/translation="MDAVQEFERHTELEIQSALOEDVIAGLNAMVAELROTLDLQQA
QLRLLYKMQDRNPDAQEYSLRDEIPPHY"
complement(3271..3423)
/gene="NMA0372"
complement(3271..3423)
/gene="NMA0372"
/note="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=1
/product="very hypothetical protein NMA0372"
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/db_xref="SPTREMBL:O9JWH3"
/translation="MEGRSLPLGSDGFFLAVVEYVPVGFICLOGKTAIFRSQGRPHRI
KGI:SD"
3562..3571
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3586..4356)
/gene="thif"
complement(3586..4356)
/gene="thif"
/note="NMA0373, thif, probable Thif protein. len: 256 aa;
similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
fasta scores: E(): 0, 43.1% identity in 246 aa overlap,
and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
protein. (249 aa), fasta scores: E(): 0, 43.9% identity
in 244 aa overlap (note that N.m. does not have orthologs of
any other molybdopterin biosynthesis proteins). Contains
pfam match to entry pf00899 Thif_family, Thif family"
/codon_start=1
/transl_table=1
/product="Thif protein"
/protein_id="CAB83674.1"

```



Result No.	Query Match	Score	Length	DB	ID	Description
1	100.0	1442	19	AAV10449	K. oxytoca R-speci	
2	49.2	945	21	AAAC64148	Enterobacter Cloac	
3	49.2	1444	21	AAAC64149	Enterobacter Cloac	
4	41.2	34980	22	AAK41224	Pyrococcus abyssi	
5	39.2	27	3200	AAQ66920	Cell death hif gBK	
6	39.2	27	3988	ABL08913	Drosophila melanog	
7	39.2	27	14867	ABL08912	Drosophila melanog	
8	39.2	27	25778	AAAB14177	N. meningitidis pa	
9	39.2	27	345980	AAAF21612	Neisseria meningit	

PA (1.0N% ) LONZA AG.  
yy

P: Brion W, Naughton A, Robins K, Shaw N, Tinschert A;

PI Zimmermann I;

XX WPI: 1998-101063/39.

DR P-PSDB: AAW40263.

XX Tri-fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation

PT - by stereoselective hydrolysis of corresponding racemic amide using

PT microorganism or derived enzyme, used as drug intermediate

XX

PS Claim 7a: Page 32-34; 68pp; German.

XX This DNA sequence encodes a R-specific amidohydrolase isolated from

CC Klebsiella oxytoca strain PRS1 which allows the microorganism to

CC utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole

CC nitrogen source. This amidohydrolase is used in a process for preparing

CC (R)-isomers of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which

CC is cheaper than prior art optical resolution of the racemate using

CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.

XX

SQ Sequence 1442 BP; 365 A; 350 C; 370 G; 337 T; 0 other;

Query Match 100.0%; Score 1442; DB 19; Length 1442;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCGGGAACCTCCATGTGCGCGTGATCCTGGTCGAGCAGGATATCGCATGATCCAGCGG 50

DB 1 CCCGGGAACCTCCATGTGCGCGTGATCCTGGTCGAGCAGGATATCGCATGATCCAGCGG 50

QY 61 CGGCACAGCGTGGCGGTAAATGGATAAAGCCCTGGTTAGAAACCTGACCCCAAC 120

DB 61 CGGCACAGCGTGGCGGTAAATGGATAAAGCCCTGGTTAGAAACCTGACCCCAAC 120

QY 121 AGCTCTCGATGATCTTTTAATGCTGCTCATCTGCTCTGTAACCTAAGCTATTAAT 180

DB 121 AGCTCTCGATGATCTTTTAATGCTGCTCATCTGCTCTGTAACCTAAGCTATTAAT 180

QY 181 AGTGGAGAATAACATATGAATGGTGGAGAATCCATTATGCCAAAGCGGTGG 240

DB 181 AGTGGAGAATAACATATGAATGGTGGAGAATCCATTATGCCAAAGCGGTGG 240

QY 241 TCCCGGGGTAAACCGGTAAACCGATCCTCAGCGAAGAATGCAAAAGATTTCATTA 300

DB 241 TCCCGGGGTAAACCGGTAAACCGATCCTCAGCGAAGAATGCAAAAGATTTCATTA 300

QY 301 CACCAATTGGCTTATTCACACCGCTGCTGACCATGGAACCGGTGACCGGATATTGT 360

DB 301 CACCAATTGGCTTATTCACACCGCTGCTGACCATGGAACCGGTGACCGGATATTGT 360

QY 361 CGACATCGAGATGCTTTTGAGGTGCTATCAATTCGACAGCATATTCGAGCGAGT 420

DB 361 CGACATCGAGATGCTTTTGAGGTGCTATCAATTCGACAGCATATTCGAGCGAGT 420

QY 421 GCTAAATATGCCCTTCTCAACCCACAAACGGACCGATATGCTCAATGCGCGGAGAA 480

DB 421 GCTAAATATGCCCTTCTCAACCCACAAACGGACCGATATGCTCAATGCGCGGAGAA 480

QY 481 AGGTGATGCTGCTGCTTATCGAATCCATGTTGCGCGCGTATGATCCCTACGG 540

DB 481 AGGTGATGCTGCTGCTTATCGAATCCATGTTGCGCGCGTATGATCCCTACGG 540

QY 541 CATCTGCCCATGATTCGGATTTTGGCGGATGACCGGACCGATGAGCGGATGCT 600

DB 541 CATCTGCCCATGATTCGGATTTTGGCGGATGACCGGACCGATGAGCGGATGCT 600

QY 601 CAATGATCGCTGCCAGAAAAGGTGCGCATGATTAAACTCCACAGTGAAGAAGTCTACTG 660

DB 601 CAATGATCGCTGCCAGAAAAGGTGCGCATGATTAAACTCCACAGTGAAGAAGTCTACTG 660

QY 661 GAGCAACCGCATATTCCTTATTAACCCCATATTTGGCACCTTGAGCGTATGGCCAGA 720

DB 661 GAGCAACCGCATATTCCTTATTAACCCCATATTTGGCACCTTGAGCGTATGGCCAGA 720

QY 721 AATTGACTCAATCAATTCACATGACCCAGACAATACAGCGGGGAATATGGAATGTCGGGA 780

DB 721 AATTGACTCAATCAATTCACATGACCCAGACAATACAGCGGGGAATATGGAATGTCGGGA 780

QY 781 TATAGNCCAGGAGATATACCIATCTCCGGTAGCTGCGCTTGAGAGCCCGCTGTTAT 840

DB 781 TATAGNCCAGGAGATATACCIATCTCCGGTAGCTGCGCTTGAGAGCCCGCTGTTAT 840

QY 841 TGGTATGCCCATGCTTGTTCAGGGTATGATGGTATGATTTGCGGGACCGCAGTAGATGTC 900

DB 841 TGGTATGCCCATGCTTGTTCAGGGTATGATGGTATGATTTGCGGGACCGCAGTAGATGTC 900

QY 901 CTCATATCACCACCATCAAGTCGATTTGATCAAGAACTGGCAGCTTTCCTGGCCACGAAT 960

DB 901 CTCATATCACCACCATCAAGTCGATTTGATCAAGAACTGGCAGCTTTCCTGGCCACGAAT 960

QY 961 GGAGAAATGCCGAAATATATGAGTATTTGGCAGTGCAGCTCGCTGGAGGATGCGACGG 1020

DB 961 GGAGAAATGCCGAAATATATGAGTATTTGGCAGTGCAGCTCGCTGGAGGATGCGACGG 1020

QY 1021 AATTGCATATCCGACTTAATTTACTGCTGGTATGAGAGACTTTGGCTTCGAACAATGGGA 1080

DB 1021 AATTGCATATCCGACTTAATTTACTGCTGGTATGAGAGACTTTGGCTTCGAACAATGGGA 1080

QY 1081 TGCCTACATGCTTCTGAGTCAATGCGCAAAAGTGGCTGGCAACATGGTGCAGCCCA 1140

DB 1081 TGCCTACATGCTTCTGAGTCAATGCGCAAAAGTGGCTGGCAACATGGTGCAGCCCA 1140

QY 1141 ATACACCGTTGGCGGATGCTGAAACAAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200

DB 1141 ATACACCGTTGGCGGATGCTGAAACAAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200

QY 1201 TGAACATACCCGGATGATAGTGGGGTAAATGCTGAGTTCACAACTCGCTATTTAA 1260

DB 1201 TGAACATACCCGGATGATAGTGGGGTAAATGCTGAGTTCACAACTCGCTATTTAA 1260

QY 1261 CAGCTAAAGCAGTGCATATGGGGCCAGATACCCCATCAATATTTGTTTACTTCTCC 1320

DB 1261 CAGCTAAAGCAGTGCATATGGGGCCAGATACCCCATCAATATTTGTTTACTTCTCC 1320

QY 1321 TTAGCGGAGTACCGGCGCACAGAGTGTGCACAAATGGCGGGAGCAACCCAGCTATT 1380

DB 1321 TTAGCGGAGTACCGGCGCACAGAGTGTGCACAAATGGCGGGAGCAACCCAGCTATT 1380

QY 1381 GCGGAAATTAATCAAAATGCGGCGATCAAGAGTGTGCACAAATGGCGGGAGCAACCCAGCTATT 1440

DB 1381 GCGGAAATTAATCAAAATGCGGCGATCAAGAGTGTGCACAAATGGCGGGAGCAACCCAGCTATT 1440

QY 1441 CC 1442

DB 1441 CC 1442

RESULT 2

AA264148

ID AA264148 standard; DNA; 945 BP.

XX AA264148;

AC AA264148;

CT 23 FEB-2001 (first entry)

XX

TF Enterobacter cloacae stereoselective amidase DNA coding sequence.

XX

KW Amidase: stereoselective hydrolysis; alpha-amino acid amide;

KK alpha-hydroxy acid amide; optically active product;

XX recombinant production; ds.

XX

OS Enterobacter cloacae.

XX

PN W0200063354-A1.

XX

PD 26-OCT-2000.

```

XX PF 17-APR-2000; 2000WO-JP02492.
XX PR 16-APR-1999; 99JP-0109328.
XX PA (MITR ) MITSUBISHI RAYON CO LTD.
XX PI Nakamura T, Yu F;
XX WP: 2000-672731/65.
XX DR P-PSDB; AAB29631.
XX PT Novel amidase gene encoding protein which stereoselectively hydrolyzes
XX PI alpha-amino-acid amides and alpha-hydroxy-acid amides, with enhanced
XX PT catalytic capability to yield optically active alpha-amino-acids and
XX PT alpha-hydroxy-acids.
XX PS Claim 3; Page 23-25; 32pp; Japanese.
XX CC The invention relates to a novel amidase, and the gene encoding the
XX CC novel amidase, from Enterobacter cloacae. The novel amidase
XX CC stereoselectively hydrolyses alpha-amino acid amides and alpha-
XX CC hydroxy acid amides to give optically active alpha-amino acids and
XX CC alpha-hydroxy acids. The invention also relates to variants of the novel
XX CC amidase which retain activity, recombinant vectors and host cells
XX CC containing the novel amidase gene, and recombinant production of the
XX CC novel amidase. Genetically manipulated microorganisms comprising the
XX CC novel amidase gene have remarkably elevated catalytic capability as
XX CC compared to those produced by conventional methods. The present sequence
XX CC represents the DNA coding sequence of the novel Enterobacter cloacae
XX CC amidase.
XX SQ Sequence 945 BP; 169 A; 285 C; 285 G; 205 T; 0 other;

Query Match 3.4%; Score 49.2; DB 21; Length 945;
Best Local Similarity 51.4%; Pred. No. 1.7e-05;
Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

CY 741 TGACGCCAGACATCAGCGCGGAATATGATGTCGCGATATAGACACGAGGATATTA 800
DB 741 TGACGCCAGACATCAGCGCGGAATATGATGTCGCGATATAGACACGAGGATATTA 800
CY 473 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 532
DB 473 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 532
CY 801 CCTATCTGCCGCTACGTCGCGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
DB 801 CCTATCTGCCGCTACGTCGCGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
CY 533 TATGGCTACCGTGGAGTAGAGGCGCGTGTCTTCTATGTCGACACCGCGCGCG 592
DB 533 TATGGCTACCGTGGAGTAGAGGCGCGTGTCTTCTATGTCGACACCGCGCGCG 592
CY 861 AGGTGATGTCGATGTCGCGGACCGCGAGTAGAGTTGCTCAATCACCACCAATCAAG 920
DB 861 AGGTGATGTCGATGTCGCGGACCGCGAGTAGAGTTGCTCAATCACCACCAATCAAG 920
CY 593 AGGCGGATGTCGAGTCTGCGGACCGCGGATTTGAAAGCGCGCGATGAGCGTGGTGAAGC 552
DB 593 AGGCGGATGTCGAGTCTGCGGACCGCGGATTTGAAAGCGCGCGATGAGCGTGGTGAAGC 552
CY 921 TCGATTGATCAAGAACTGCGAGCTTTCCTGGCCGCAATGG 962
DB 921 TCGATTGATCAAGAACTGCGAGCTTTCCTGGCCGCAATGG 962
CY 653 TGGAGTGGTTAAGATA:GCCGCTGAAGACGCGCGTTCG 694
DB 653 TGGAGTGGTTAAGATA:GCCGCTGAAGACGCGCGTTCG 694

RESULT 3
AAC64149
ID AAC64149 standard; DNA; 1444 BP.
XX AC AAC64149;
XX DT 23-FEB-2001 (first entry)
XX DE Enterobacter cloacae stereoselective amidase DNA.
XX KW Amidase; stereoselective hydrolysis; alpha-amino acid amide;
XX KW alpha-hydroxy acid amide; optically active product;
XX KW recombinant production; ds.
XX OS Enterobacter cloacae.
XX PN W0200063354-A1.
XX FT

```

```

PD 26-OCT-2000.
XX PF 17-APR-2000; 2000WO-JP02492.
XX PR 16-APR-1999; 99JP-0109328.
XX PA (MITR ) MITSUBISHI RAYON CO LTD.
XX PI Nakamura T, Yu F;
XX WP: 2000-672731/65.
XX DR P-PSDB; AAB29631.
XX PT Novel amidase gene encoding protein which stereoselectively hydrolyzes
XX PI alpha-amino-acid amides and alpha-hydroxy-acid amides, with enhanced
XX PT catalytic capability to yield optically active alpha-amino-acids and
XX PT alpha-hydroxy-acids.
XX PS Example 1; Page 26-29; 32pp; Japanese.
XX CC The invention relates to a novel amidase, and the gene encoding the
XX CC novel amidase, from Enterobacter cloacae. The novel amidase
XX CC stereoselectively hydrolyses alpha-amino acid amides and alpha-
XX CC hydroxy acid amides to give optically active alpha-amino acids and
XX CC alpha-hydroxy acids. The invention also relates to variants of the novel
XX CC amidase which retain activity, recombinant vectors and host cells
XX CC containing the novel amidase gene, and recombinant production of the
XX CC novel amidase. Genetically manipulated microorganisms comprising the
XX CC novel amidase gene have remarkably elevated catalytic capability as
XX CC compared to those produced by conventional methods. The present sequence
XX CC represents the DNA encoding the novel Enterobacter cloacae amidase.
XX SQ Sequence 1444 BP; 270 A; 422 C; 450 G; 302 T; 0 other;

Query Match 3.4%; Score 49.2; DB 21; Length 1444;
Best Local Similarity 51.4%; Pred. No. 2.3e-05;
Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

CY 741 TGACGCCAGACATCAGCGCGGAATATGATGTCGCGGATATAGACACGAGGATATTA 800
DB 741 TGACGCCAGACATCAGCGCGGAATATGATGTCGCGGATATAGACACGAGGATATTA 800
CY 833 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 892
DB 833 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 892
CY 801 CCTATCTGCCGCTACGTCGCGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
DB 801 CCTATCTGCCGCTACGTCGCGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
CY 833 TATGGCTACCGTGGAGTAGAGGCGCGTGTCTTCTATGTCGACACCGCGCGCG 952
DB 833 TATGGCTACCGTGGAGTAGAGGCGCGTGTCTTCTATGTCGACACCGCGCGCG 952
CY 861 AGGTGATGTCGATGTCGCGGACCGCGAGTAGAGTTGCTCAATCACCACCAATCAAG 920
DB 861 AGGTGATGTCGATGTCGCGGACCGCGAGTAGAGTTGCTCAATCACCACCAATCAAG 920
CY 933 AGGCGGATGTCGAGTCTGCGGACCGCGGATTTGAAAGCGCGCGATGAGCGTGGTGAAGC 1012
DB 933 AGGCGGATGTCGAGTCTGCGGACCGCGGATTTGAAAGCGCGCGATGAGCGTGGTGAAGC 1012
CY 921 TCGATTGATCAAGAACTGCGAGCTTTCCTGGCCGCAATGG 962
DB 921 TCGATTGATCAAGAACTGCGAGCTTTCCTGGCCGCAATGG 962
CY 1013 TGGAGTGGTTAAGATA:GCCGCTGAAGACGCGCGTTCG 1054
DB 1013 TGGAGTGGTTAAGATA:GCCGCTGAAGACGCGCGTTCG 1054

RESULT 4
AAH41224
ID AAH41224 standard; DNA; 349980 BP.
XX AC AAH41224;
XX DT 29-OCT-2001 (first entry)
XX DE Pyrococcus abyssi genomic fragment #3.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX KW Pyrococcus abyssi.
XX OS Pyrococcus abyssi.
XX FT Key Location/Qualifiers
XX FT misc_feature 1..49980
XX FT /tag- a
XX FT /note- *This sequence overlaps with the 3' end of

```



```

FT misc_feature AAH41223
FT 300001..349980
FT /tag= b
FT /note= "This sequence overlaps with the 5' end of
FT AAH41225"
XX
XX FR2792651-A1.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999: 99PR-3005034.
XX
XX 21-APR-1999: 99PR-0005034.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prier D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
XX WPI: 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry .
XX
XX Claim 1: Page 347-443; 1657pp: French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P.
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41223
XX and the 3' end of this sequence overlaps with the 5' end of AAH41225. The
XX proteins of the present invention have various potential industrial uses,
XX since the proteins are stable at very high temperatures, some up to 110
XX degrees centigrade.
XX Note: This patent is in the same patent family as WC2000065052, which
XX contains additional sequences as shown in AAB99132-AAB99143.
XX AAH75903-AAH75920 and AAG56436.
XX
XX Sequence 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 other;
SQ

```

Query Match 2.9%; Score 41.2; DB 22; Length 349980;  
 Best Local Similarity 53.8%; Pred. No. 0.39;  
 Matches 85; Conservative 0; Mismatches 73; Indels 6; Gaps 6;

```

QY 743 ACCCCAGCAATCACGCGCGGAATATGGATGTCGCGATATAGGACCGAGGATATTACC 802
DQ 282390 ACCCCAGGAACACGCGCGGAACATGGACACAAACCTAATCAGAAAGGACACCAAT 282449
QY 803 TATCTCCGGTACGTCGCGCTGGAGCCCGCTTTTATTTGTTGATGTCGCTGTCAG 862
DQ 282450 TACTTCCCTGCTCTCTTATGAGAGCTTACCTCGGCTAGGCGATTTCGATGCCGTAATG 282509
QY 863 GGTGATGGTGAGATTTCGCGGACCGCAGTAGAGATTTC 900
DQ 282510 GGAGATCGCGAGCTTCGCTCTCGCTCTGAGGTTTC 282547

```

RESULT 5  
 AAQ66950/c  
 ID AAQ66950 standard; cDNA; 3900 BP.  
 XX  
 AC AAQ66950;  
 XX  
 XX 15-FEB-1995 (first entry);  
 XX  
 DE Cell death hid cDNA.  
 XX  
 XX Cell death; head involution defective; hid gene; reaper gene; rpr;  
 XX apoptosis; transgenic animal; antisense; ss.  
 XX  
 OS *Drosophila melanogaster*.

```

XX FH Key Location/Qualifiers
XX FT 450..1632
XX FT /*tag= a
XX
XX DN W05416071-A.
XX
XX XX 21-JUL-1994.
XX
XX XX 14-JAN-1994; 94MO-US00500.
XX
XX XX 15-JAN-1993; 93US-0004957.
XX XX 17-SEP-1993; 93US-0123343.
XX
XX (MAST ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Abrams JM, Grether ME, Steller H, White K;
XX
XX WPI: 1994-249218/30.
XX P-PSDB; AAR55791.
XX
XX Isolated cell death genes from Drosophila - and novel assays for
XX apoptotic cell deaths and apoptotic and necrotic cell deaths,
XX based on selective staining by toluidine blue, acridine orange
XX and Nile blue
XX
XX Disclosure: Fig 6; 93pp; English.
XX
XX CC The reaper (rpr) and head involution defective (hid) genes, mapping
XX to position 75c1.2 of D. melanogaster chromosome 3, exhibit
XX expression patterns related to the pattern of cell death during
XX Drosophila embryogenesis. Cell death genes, or antisense sequences,
XX can be used to reduce or abolish apoptosis, e.g. in transgenic
XX animals. An hid cDNA sequence is given in AAQ66950, and the
XX deduced sequence of the encoded HID protein in AAR55791.
XX
XX SQ Sequence 3900 BP; 1288 A; 876 C; 736 G; 1000 T; 0 other;

```

Query Match 2.7%; Score 39.2; DB 15; Length 3900;  
 Best Local Similarity 49.1%; Pred. No. 0.11;  
 Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

```

QY 320 ACACCGCTCTGACCATCGAACCGGTGACCGGATATATGTCGACATCGAGATGCTTT 379
DQ 1458 ACCACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1399
QY 380 GAAGGTGCTATCAATTCGCAACAGGATATTCGAGCCAGTTCGTAATAATGCCCTTCTC 439
DQ 1398 GGAGCTCTCTTCTTCGCGGGGGGAGTGGTCCATGTAATCTCGACACGCTCCGGCC 1339
QY 440 AACCCACAAACGAGCGATCATGTGTAATGCGCGGAGAAAGGTGATGTCGCTGCTC 499
DQ 1338 GGCGGTAAAGTTCGCTAGCGATCGCAAACTCGTCCCAAGTGGGTCATGATCGCTCTG 1279
QY 500 TATATGAATCCATGTTGCGCGCGCGCGGTGA 531
DQ 1276 G-TACTCGCTCATCTCTCGCGCGCGGTGCA 1247

```

RESULT 6  
 ABL08913/c  
 ID ABL08913 standard; cDNA; 3988 BP.  
 XX  
 AC ABL08913;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 21221.  
 XX  
 XX *Drosophila*; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical; gene; ss.  
 XX  
 OS *Drosophila melanogaster*.

```
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-655860/75.
XX P-PSDB; ABB64810.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1: SEQ ID NO 21221; 21pp - Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 3988 BP; 1316 A; 893 C; 753 G; 1026 T; 0 other;
XX
XX Query Match 2.7%; Score 39.2; DB 23; Length 19867;
XX Best Local Similarity 49.1%; Pred. No. 0.3;
XX Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
XX
XX QY 320 ACACCCGCTCTGACCATCGAACCCGGTGACCGGATTATTGTCGACACTCGAGATGCTTT 379
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 1545 ACCACCGGGCIGGCTATCGGTATGCGACACCGGATTATTGCTGCTCGAGTGGCTATT 1486
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCCAGTGTCTTAAAAATGCCCTTTTC 439
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 1485 GGAGCTCTTCTTTCGGGGGGGGATCGCTCCATTGAACCTCTCGACAGCTCCCGGCC 1426
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX QY 440 AACCCACAAAACGGACCATCATGTGTAATGGCGGGAGAAAGTGTATGTCGCTGTC 499
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 1425 GGCGGTAAGTGTGCTAGCGATCGCAAACTCGTCCCAAGTGGCTCATGATCGCTCIG 1366
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX QY 500 TATATCGAATCATGTTGCCCGCGCGGCTTGA 531
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 1365 GTACTCGGCTCATCTCGCGCGTTCGAGGTCA 1334
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX RESULT 7
XX ABL06912/c
XX ID ABL06912 standard; cDNA; 19867 BP.
XX AC ABL06912;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21218.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX OS Drosophila melanogaster.
XX
```

```
PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-655860/75.
XX P-PSDB; ABB64809.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1: SEQ ID NO 21218; 21pp - Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 19867 BP; 5006 A; 3967 C; 3859 G; 6035 T; 0 other;
XX
XX Query Match 2.7%; Score 39.2; DB 23; Length 19867;
XX Best Local Similarity 49.1%; Pred. No. 0.3;
XX Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
XX
XX QY 320 ACACCCGCTCTGACCATCGAACCCGGTGACCGGATTATTGTCGACACTCGAGATGCTTT 379
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 14456 ACCACCGGGCIGGCTATCGGTATGCGACACTCGGATTATTGCTGCTCGAGTGGCTATT 14397
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCCAGTGTCTTAAAAATGCCCTTTTC 439
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 14396 GGAGCTCTTCTTTCGGGGGGGGATCGCTCCATTGAACCTCTCGACAGCTCCCGGCC 14337
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX QY 440 AACCCACAAAACGGACCATCATGTGTAATGGCGGGAGAAAGTGTATGTCGCTGTC 499
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 14336 GGCGGTAAGTGTGCTAGCGATCGCAAACTCGTCCCAAGTGGCTCATGATCGCTCIG 14277
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX QY 500 TATATCGAATCATGTTGCCCGCGCGGCTTGA 531
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 14276 GTACTCGGCTCATCTCGCGCGTTCGAGGTCA 14245
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX RESULT 8
XX ABL06912/c
XX ID ABL06912 standard; DNA; 26778 BP.
XX AC ABL06912;
XX
XX DT 04-DEC-2000 (first entry)
XX
XX DE N. meningitidis partial DNA sequence gnm_25 SEQ ID NO:25.
XX
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX Meningococcus B; MenB; ds.
XX
XX OS Neisseria meningitidis.
XX
```

```
FN WO200022430-A2.
PD 20-APR-2000.
XX 06-OCT-1999; 99WO-US23573.
XX 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX (CHIR ) CHIRON CORP.
PA
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
PI Rappuoli R, Pizza M;
DR WPI: 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PI used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea .
XX
XX Claim 7; Page 524-531; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAB81453 to AAB82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAB81260 to AAB81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAB81254 to
CC AAB81259 and AAB81304 to AAB81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAB81322 to
CC AAB81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
XX Sequence 26778 BP; 6602 A; 7308 C; 5846 G; 6313 T; 4 other;
XX
XX Query Match 2.7%; Score 39.2; DB 21; Length 26778;
XX Best Local Similarity 52.4%; Pred. No. 0.37;
XX Matches 86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
XX
XX 188 GAATAACATATGAATGCTTGAGAGATCCATTATGCGCAACCGCGTGTGTGCGCGG 247
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 23055 GCATCTTAAACAAAACACGCAAAAGCTATGATGACCGGAAGCGAGTTTATCCGCGG 22994
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 248 CGTAAACCGGTAAAGCATCATCTGACGAGAGAAATGCAAAAGAGTTTCATACACCAT 307
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 22995 AGCGAGGCAATTATTGACACATCGAACACCAATATCGGNAACGCGCTGGGATTTCGAC 22936
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 308 GGCCCTTATTCACACCGTCTGACCATCGACATCGACCGCGTGACCG 351
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 22935 TGCCGGTTTCCGGAACAGTCTTCGACATCGAAGCGGAGACGG 22892
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX RESULT 9
XX AAF21612
XX ID AAF21612 standard; DNA: 349960 BP.
XX
XX AC AAF21612;
```

```
XX 13-MAR-2001 (first entry)
XX
XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
XX diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX ds.
XX
XX Neisseria meningitidis.
XX
XX WO2000066791-A1.
XX 09-NOV-2000.
XX
XX 05-MAR-2000; 2000WO-US05928.
XX
XX 30-APR-1999; 99US-0132068.
PR 06-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V,
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R,
PI Frazer CM, Grandi G;
XX
XX WPI: 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PI frames are used to detect, treat and prevent Neisserial infections .
XX
XX Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
XX Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;
XX
XX Query Match 2.7%; Score 39.2; DB 21; Length 349980;
XX Best Local Similarity 52.4%; Pred. No. 1.9;
XX Matches 86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
XX
XX 189 GAATAACATATGAATGCTTGAGAGATCCATTATGCGCAACCGCGTGTGTGCGCGG 247
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 27593 GCATCTTAAACAAAACACGCAAAAGCTATGATGACCGGAAGCGAGTTTATCCGCGG 276040
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 246 CGTAAACCGGTAAAGCATCATCTGACGAGAGAAATGCAAAAGAGTTTCATACACCAT 307
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 276041 AGCGAGGCAATTATTGACACATCGAACACCAATATCGGNAACGCGCTGGGATTTCGAC 276100
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX
XX 308 GGCCCTTATTCACACCGTCTGACCATCGAAGCGGAGACCG 351
```





CC thought to inhibit Hid induced cell death. Mutant forms of the hid  
CC gene, hidA1a3 and hidA1a5 have been generated in which DNA encoding  
CC phosphoacceptor residues is replaced with DNA encoding  
CC non-phosphorylatable amino acids. These mutants are not suppressed  
CC by MAPK induced phosphorylation. Compositions from the mutant gene  
CC are useful in screening assays for antagonists and agonists of Hid  
CC phosphorylation by mitogen activated protein kinase (MAPK). They may  
CC also be used to identify new constituents of the Hid signalling  
CC pathway. The new compositions may be useful in identifying new homologs  
CC of Hid or its natural mutations. The present sequence is the cDNA  
CC of the hidA1a3 mutant.  
XX  
SQ Sequence 1233 BP; 247 A; 453 C; 330 G; 203 T; 0 other;  
  
Query Match 2.6%; Score 37.6; DB 21; Length 1233;  
Best Local Similarity 48.6%; Pred. No. 0.18;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
  
QY 320 ACACCGCTCTGACCATCGAACCCGGTACCGGATTATTGTGACACATCGAGATGCTTT 373  
DB 1059 ACCACCGGACTGGCTGCGGTATGCGAGACTGAGTATTGCTGCTGAGTGGCTATT 1005  
  
QY 380 GAAGTGTCTATCAATTCGGAACAGATATTCGACCCAGTTCCTAATAATGCCCTTCTC 439  
DB 999 GGAGCTCTCTCTCTCCGGCGGGGATGCTGCAATTGAACCTCTGACAGACCTCCCGGCC 940  
  
QY 440 AACCCCAAAACGGACCGATCATGTCATATGCGCGGAGAAAGGTGATGCTGCTGCTGC 499  
DB 939 GSCCGTAAAGTGTGCTATGATGCGCAACTGCTCCCAAGTGGCTGATGCTGCTG 890  
  
QY 500 TATATCGAATCCATGTCGCCCGCGCGCTTGA 531  
DB 879 GTACTCGGCTCACTCTCCCGCGTGGAGGTCA 843  
  
RESULT 14  
AAA27104/C  
ID AAA27104 standard; cDNA; 1233 BP.  
XX  
AC AAA27104;  
XX  
31-JUL-2000 (first entry)  
XX  
XX Drosophila melanogaster hidA1a5 cDNA.  
XX  
XX Fruit fly; Hid; mitogen activated protein kinase p42/p44;  
KW MAPK; phosphorylation; apoptosis; ss.  
XX  
XX Drosophila melanogaster.  
OS Synthetic  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1233  
FT FT /\*tag= a  
FT FT /product= HidA1a5 protein  
FT FT 361..362  
FT FT /\*tag= b  
FT FT /label= wild type A G replaced with G C  
FT FT 442  
FT FT /\*tag= c  
FT FT /label= Wild type A replaced with G  
FT FT 536  
FT FT /\*tag= d  
FT FT /label= Wild type A replaced with G  
FT FT 682  
FT FT /\*tag= e  
FT FT /label= Wild type A replaced with G  
FT FT 751..752  
FT FT /\*tag= f  
FT FT /label= wild type A G replaced with G C  
XX  
XX WO200026634-A2.  
XX  
XX

PD 11-MAY-2000.  
XX  
XX 29-OCT-1999; 99WO-US25514.  
XX  
XX 29-OCT-1998; 98US-0106108.  
PR 29-OCT-1999; 99US-0431573.  
XX  
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
XX Steller H, Agapite J, Bergmann A, McCall K;  
XX  
XX WP1: 2000-376155/32.  
XX  
XX P-253; RAY94424.  
XX  
XX Composition, useful in screening assays for antagonists and agonists of  
XX Hid phosphorylation by mitogen activated protein kinase (MAPK).  
XX comprises DNA having an oligonucleotide sequence of mutant forms of the  
XX Drosophila hid gene -  
XX  
XX Claim 1: Fig 11; 58pp; English.  
XX  
XX The product of the Drosophila melanogaster hid gene is responsible for  
XX inducing apoptosis in embryos. Activation of the Ras/MAPK pathway is  
XX thought to inhibit Hid induced cell death. Mutant forms of the hid  
XX gene, hidA1a3 and hidA1a5 have been generated in which DNA encoding  
XX phosphoacceptor residues is replaced with DNA encoding  
XX non-phosphorylatable amino acids. These mutants are not suppressed  
XX by MAPK induced phosphorylation. Compositions from the mutant gene  
XX are useful in screening assays for antagonists and agonists of Hid  
XX phosphorylation by mitogen activated protein kinase (MAPK). They may  
XX also be used to identify new constituents of the Hid signalling  
XX pathway. The new compositions may be useful in identifying new homologs  
XX of Hid or its natural mutations. The present sequence is the cDNA  
XX of the hidA1a5 mutant.  
XX  
SQ Sequence 1233 BP; 245 A; 454 C; 331 G; 203 T; 0 other;  
  
Query Match 2.6%; Score 37.6; DB 21; Length 1233;  
Best Local Similarity 48.6%; Pred. No. 0.18;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
  
QY 320 ACACCGCTCTGACCATCGAACCCGGTACCGGATTATTGTGACACATCGAGATGCTTT 379  
DB 1059 ACCACCGGACTGGCTGCGGTATGCGAGACTGAGTATTGCTGCTGAGTGGCTATT 1000  
  
QY 380 GAAGTGTCTATCAATTCGGAACAGATATTCGACCCAGTTCCTAATAATGCCCTTCTC 439  
DB 999 GGAGCTCTCTCTCTCCGGCGGGGATGCTGCAATTGAACCTCTGACAGACCTCCCGGCC 940  
  
QY 440 AACCCCAAAACGGACCGATCATGTCATATGCGCGGAGAAAGGTGATGCTGCTGCTGC 499  
DB 939 GSCCGTAAAGTGTGCTATGATGCGCAACTGCTCCCAAGTGGCTGATGCTGCTG 880  
  
QY 500 TATATCGAATCCATGTCGCCCGCGCGCTTGA 531  
DB 879 GTACTCGGCTCACTCTCCCGCGTGGAGGTCA 848  
  
RESULT 15  
ABL20343  
ID ABL20343 standard; DNA; 4810 BP.  
XX  
XX ABL20343;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12502.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
OS  
XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw mode:

Run on: July 6, 2003, 04:47:45 ; Search time 97 Seconds  
(without alignments)  
4559.050 Million: cell updates/soc

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgagcc.....aatgaattcattggatcc 1442

Scoring table: IDENTITY\_NJC

Gapop 10.0 , Gap\*xt 1.0

Searched: 441462 seqs, 153338361 residues

Total number of hits satisfying chosen parameters: 38272;

Minimum DB seq length: 0

Maximum DB seq length: 2060000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

1: /cgn2\_6/plodata/1/ina/5A.COMB.seq.\*

2: /cgn2\_6/plodata/1/ina/5B.COMB.seq.\*

3: /cgn2\_6/plodata/1/ina/5A.COMB.seq.\*

4: /cgn2\_6/plodata/1/ina/65.COMB.seq.\*

5: /cgn2\_6/plodata/1/ina/PCITUS.COMB.seq.\*

6: /cgn2\_6/plodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37.5	2.5	501	1	US-08-123-343A-15
C 2	37.6	2.5	1233	4	US-09-431-573-1
C 3	37.6	2.6	1233	4	US-09-431-573-2
C 4	37.6	2.6	1233	4	US-09-431-573-3
C 5	37.6	2.6	3500	1	US-08-123-343A-6
C 6	35.2	2.4	428	4	US-09-144-367-7
C 7	34.2	2.4	5733	2	US-08-473-558A-1
C 8	33.2	2.3	3201	2	US-08-633-770A-3
C 9	33.2	2.3	4726	1	US-06-633-770A-11
C 10	32.8	2.3	7218	1	US-08-232-463-14
C 11	32.4	2.2	3376	4	US-09-309-572-3
C 12	32.4	2.2	9551	1	US-08-056-200-93
C 13	32.4	2.2	9551	2	US-08-800-644-93
C 14	32.2	2.2	1226	1	US-08-374-688-4
C 15	32.2	2.2	1349	1	US-08-374-688-1
C 16	32.2	2.2	6254	4	US-08-927-219-126
C 17	31.8	2.2	1817	4	US-08-943-731-193
C 18	31.8	2.2	20084	4	US-08-943-731-5
C 19	31.6	2.2	1380	4	US-09-513-783A-169
C 20	31.2	2.2	2167	2	US-08-743-637B-16
C 21	31.2	2.2	2167	3	US-08-743-637B-16
C 22	31	2.1	4403765	4	US-09-103-840A-2
C 23	31	2.1	441529	4	US-09-103-840A-1
C 24	30.8	2.1	1187	5	PCR-US95-13536-2
C 25	30.8	2.1	1188	5	PCR-US95-13536-1
C 26	30.6	2.1	1542	4	US-09-202-894B-3
C 27	30.6	2.1	4403765	4	US-09-103-840A-2

C 28	30.6	2.1	4411529	4	US-09-103-840A-1
C 29	30.4	2.1	1049	3	US-08-682-643-1
C 30	30.2	2.1	747	4	US-08-998-416-1016
C 31	30.2	2.1	10095	3	US-08-822-586-45
C 32	30	2.1	1614	4	US-09-522-217-7
C 33	30	2.1	2665	3	US-09-040-005-1
C 34	29.8	2.1	289	4	US-09-007-005-17
C 35	29.8	2.1	289	4	US-09-244-796-17
C 36	29.8	2.1	417	4	US-08-936-165A-17
C 37	29.4	2.0	1888	1	US-08-485-107-1
C 38	29.4	2.0	7902	4	US-08-961-527-112
C 39	29.2	2.0	1524	4	US-09-134-001C-2055
C 40	29.2	2.0	46899	1	US-08-471-119A-1
C 41	29	2.0	300	1	US-08-083-946-6
C 42	29	2.0	300	3	US-08-452-915-6
C 43	29	2.0	615	4	US-09-134-001C-138
C 44	29	2.0	850	1	US-07-611-528A-1
C 45	29	2.0	850	1	US-08-083-946-1

#### ALIGNMENTS

RESULT 1  
: Sequence 15, Application US/08123343A  
: Patent No. 5593879  
: GENERAL INFORMATION:  
: APPLICANT: Steller, Hermann  
: APPLICANT: Abrams, John M.  
: APPLICANT: Grether, Megan F.  
: APPLICANT: White, Kristin  
: TITLE OF INVENTION: Cell Death Genes of Drosophila  
: TITLE OF INVENTION: Melanogaster and Vertebrate Analogs  
: NUMBER OF SEQUENCES: 16  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
: STREET: Two Militia Drive  
: CITY: Lexington  
: STATE: MA  
: COUNTRY: US  
: ZIP: 02173  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/123,343A  
: FILING DATE: 17-SEP-1993  
: CLASSIFICATION: 80C  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/004,957  
: FILING DATE: 15-JAN-1993  
: ATTORNEY/AGENCY INFORMATION:  
: NAME: Granahan, Patricia  
: REGISTRATION NUMBER: 32,227  
: REFERENCE/DOCKET NUMBER: MIT-5907A  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 617-861-6240  
: TELEFAX: 61861-9540  
: INFORMATION FOR SEQ ID NO: 15:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 503 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: unknown  
: TOPOLOGY: unknown  
: MOLECULE TYPE: DNA (genomic)  
US-08-123-343A-15

Query Match 2.6%; Score 37.6; DB 1; Length 503;  
Best local similarity 48.6%; Pred. No. 0.02;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;





Db 939 GCGCGTAAAGTGTCTGTAAGCATGCCAAACCTCTCCCAAGTGGCTTCATGATGCTCTG 880  
Qy 500 TATATCGAATCCATGTTGCCCGCGGGCTTCA 531  
Db 879 GTACTCGGCTCATCCCGCGCGTGGAGGTCA 848

RESULT 5  
US-08-123-343A-6/c  
: Sequence 6, Application US/08123343A  
: Patent No. 5593879  
: GENERAL INFORMATION:  
: APPLICANT: Steller, Herman;  
: APPLICANT: Abrams, John M.  
: APPLICANT: Grether, Megan E.  
: APPLICANT: White, Kristin  
: TITLE OF INVENTION: Cell Death Genes of Drosophila  
: TITLE OF INVENTION: Melanogaster and Vertebrate Analogs  
: NUMBER OF SEQUENCES: 16  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
: STREET: Two Militia Drive  
: CITY: Lexington  
: STATE: MA  
: COUNTRY: US  
: ZIP: 02173  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/123.343A  
: FILING DATE: 17-SEP-1993  
: CLASSIFICATION: 50C  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/004,957  
: FILING DATE: 15-JAN-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Granahan, Patricia  
: REGISTRATION NUMBER: 32,227  
: REFERENCE/DOCKET NUMBER: MIT-5907A  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 617-861-6240  
: TELEFAX: 61861-9540  
: INFORMATION FOR SEQ ID NO: 6:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3900 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: unknown  
: TOPOLOGY: unknown  
: MOLECULE TYPE: cDNA  
US-08-123-343A-6

Query Match 2.6%; Score 37.6; DB 1; Length 3900;  
Best Local Similarity 48.68; Pred. No. 0.072;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 320 ACACCCGCTCGACATCGAACCCGGTACCGGATTTTTCGACACTCGAGATGTTT 379  
Db 1458 ACCACCGGACTGGCTGCGGTATGGCAGACCTGGATTTTTCGCTGCTCGAGTGGCTATT 1399

Qy 380 GAAGGTGCTATCAATTCGGACACAGATATCCGAGCCAGTTGCTAAAATGGCCCTTTC 439  
Db 1398 GGAGCTCTCTCTTCGCGGCGGCGGATCGCGTCCATTGAACCTCCCTGACAGCGCTCCCGCC 1339

Qy 440 AACCCACAAACGGACCGATCGTCAATGGCCGGAGAAAGTGAATGTCGCTGC 499  
Db 1338 GCGCGTAAAGTTGCTAGCGATCGCCAAACCTCTGCCAAGTGGCTTCATGATCGCTG 1279

Qy 500 TATATCGAATCCATGTTGCCCGCGGGCTTCA 531  
Db 1278 GTACTCGGCTCATCCCTCGCGCGTGGAGGTCA 1247

RESULT 6  
US-09-144-367-7/c  
: Sequence 7, Application US/09144367  
: Patent No. 6432639  
: GENERAL INFORMATION:  
: APPLICANT: Lichter, Jay  
: APPLICANT: Guido, Marco  
: TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
: FILE REFERENCE: SEQ-12P  
: CURRENT APPLICATION NUMBER: US/09/144,367  
: CURRENT FILING DATE: 1998-08-31  
: PRIOR APPLICATION NUMBER: 60/058,612  
: PRIOR FILING DATE: 1997-09-10  
: NUMBER OF SEQ ID NOS: 58  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO: 7  
: LENGTH: 429  
: TYPE: DNA  
: ORGANISM: H. sapiens  
: FEATURE:  
: NAME/KEY: Other  
: LOCATION: (0)...(0)  
US-09-144-367-7

Query Match 2.4%; Score 35.2; DB 4; Length 429;  
Best Local Similarity 57.5%; Pred. No. 0.12;  
Matches 61; Conservative 1; Mismatches 44; Indels 0; Gaps 0;

Qy 1164 ACAAAACCGTGTAGTTTACTAGGATACTAACCGGTGAACATTACCCGGATGTAGATC 1223  
Db 315 ACACATATCTCAATGTACTACAAATCACTGACTGATATTTTAAGTGGATGAATAC 256

Qy 1224 GCGTAATGTGAATGTCAAAATCGCTATTTTTAACAGCTAAAG 1269  
Db 255 ATGGCATTTATATCTCATATAAGCAGTTATTTTAAGAGAGMAAG 210

RESULT 7  
US-09-473-553A-1  
: Sequence 1, Application US/08473553A  
: Patent No. 5859338  
: GENERAL INFORMATION:  
: APPLICANT: Meyerowitz, Elliot M.  
: APPLICANT: Clark, Steven E.  
: APPLICANT: Williams, Robert W.  
: TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
: TITLE OF INVENTION: Transformed Plants, and Proteins  
: NUMBER OF SEQUENCES: 11  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Fleiss, Honbach, Test, Albritton & Herbert  
: STREET: Four Embarcadero Center, Suite 3400  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: United States  
: ZIP: 94111-4187  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/473,553A  
: FILING DATE: 06-JUN-1995  
: CLASSIFICATION: 800  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Silva, Robin M.  
: REGISTRATION NUMBER: 38,304  
: REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (415) 781-1989  
: TELEFAX: (415) 398-3249

TELEX: 910 277299  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 5733 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: unknown  
: TOPOLOGY: unknown  
: MOLECULE TYPE: DNA (genomic)  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 2434..5037  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 5117..5467  
US-08-473-553A-1

Query Match 2.4%; Score 34.2; DB 2: Length 5733;  
Best Local Similarity 50.3%; Pred. No. 1.2;  
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
Oy 1136 CCCAATACACCGTTCGGCGCAAGCTGAACAAAACCTGTTAGTTAGTAGAATAACTA 1155  
Db 3709 CTCACGGGACAGTTCGGCGGGGGCTTTCATCTACCTACCGTTCAGTTACGATATCGAACATC 3748  
Oy 1196 ACCGGTGAACATACCCGATGTAGTCGGGTAAAGTGAAGTTCAACAAACGCTATT 1255  
Db 3769 ACTGATAATTCTCTCCGCTGAACITCCGGTAACGATGTCGGCGGATGTCCTCGATCAG 3828  
Oy 1256 TTTAAGAGCTAAGCAGGTGCATATGCGGCGCAGATACACCCATCAAT 1302  
Db 3829 ATTACTCTCTAACAACTGTTTTCGGGCGAGATTCACCTGGCAT 3875

## RESULT 8

US-08-633-770A-3/c  
: Sequence 3, Application US/08633770A  
: Patent No. 5908760  
: GENERAL INFORMATION:  
: APPLICANT: Hojsten, Kirsten  
: APPLICANT: Yu, Shukun  
: APPLICANT: Kragh, Karsten  
: APPLICANT: Christensen, Tove  
: APPLICANT: Marcussen, Jan  
: TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
: TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Knobbe, Martens, Olson & Bear  
: STREET: 620 Newport Center Drive 16th Floor  
: CITY: Newport Beach  
: STATE: CA  
: COUNTRY: U.S.A.  
: ZIP: 92660  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/633,770A  
: FILING DATE: July 8, 1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/EP94/03398  
: FILING DATE: OCT-15-1994  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Altman, Daniel E  
: REGISTRATION NUMBER: 34,115  
: REFERENCE/DOCKET NUMBER: DY006.001APC  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 714-760-0404  
: TELEFAX: 714-760-9502  
: TELEX:

: INFORMATION FOR SEQ ID NO: 3:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3201 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
US-08-633-770A-3

Query Match 2.3%; Score 33.2; DB 2: Length 3201;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
Oy 243 CCGCGGTAAACGGTAAACGATCACTGACCGAAGAAATGCAAAAAGAGTTTCATTACA 302  
Db 1628 CTGGGCTTCACTTTGACAGTCACACAGACCTGTATAAGTCGTAACATTTCCGGGACC 969  
Oy 303 CCATGGCCCTTATTCACACACCGCTCTGACCATCGAACCCGGTGACCGGATTAATGTCG 362  
Db 968 GTATCGGCTACTGATACCGTAACAAATCGATACCACTACCTGCTCCAGCTTGATGAG 909  
Oy 363 ACACTCGAGATGCTTTGAGGTGGCTATCAATTCGAACAGGATAT 408  
Db 908 CCTCAGTTGGCTCTCCAAAGTCGATGGCAATTTGAGAGTASTTAT 863

## RESULT 9

US-08-633-770A-11/c  
: Sequence 11, Application US/08633770A  
: Patent No. 5908760  
: GENERAL INFORMATION:  
: APPLICANT: Hojsten, Kirsten  
: APPLICANT: Yu, Shukun  
: APPLICANT: Kragh, Karsten  
: APPLICANT: Christensen, Tove  
: APPLICANT: Marcussen, Jan  
: TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
: TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Knobbe, Martens, Olson & Bear  
: STREET: 620 Newport Center Drive 16th Floor  
: CITY: Newport Beach  
: STATE: CA  
: COUNTRY: U.S.A.  
: ZIP: 92660  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/633,770A  
: FILING DATE: July 8, 1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/EP94/03398  
: FILING DATE: OCT-15-1994  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Altman, Daniel E  
: REGISTRATION NUMBER: 34,115  
: REFERENCE/DOCKET NUMBER: DY006.001APC  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 714-760-0404  
: TELEFAX: 714-760-9502  
: TELEX:  
: INFORMATION FOR SEQ ID NO: 11:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 4726 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA

US-08-633-770A-11

Query Match 2.38; Score 33.2; DB 2; Length 4726;  
Best Local Similarity 50.08;  
Matches 83; Conservative 0; Mismatches 83; Index's 0;  
Pred. No. 2.4;

Oy	243	CCGGGGGTAAACCGGTAAACCATCC	GACGGGAGNAA	TGCAAAAGAGT	TTGATTACA	302
Db	1741	CTGGGCTTCACTTTGACCTGCCACAGACCTGTATAAAGTCGTACAAT	TCGGGAGCC	1682		
Oy	303	CNATGGCCCTTATTCACACCCCGTCTGACCATCGAACCCGGTGACCGGATATATG	TCG	362		
Db	1681	GTATCGCACTGATACCGTAACCAATCGATACCAACCATACCTGGTCCGACGTTGATG	TCAG	1622		
Oy	363	ACACTCGAGATGCTTTTGAAGGTCGTATCAATTCCGGAACAGGATAT	408			
Db	1621	CTGAGATTGCTTTTCCAAAGTCGATGGCAATTTGAGAGTAGTAT	1576			

RESULT 10

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367

: GENERAL INFORMATION:  
 : APPLICANT: CORNER, F.  
 : APPLICANT: SCHEIDLINGER, F.  
 : APPLICANT: FALKNER, F. G.  
 : TITLE OF INVENTION: RECOMBINANT FOMULPOX VIRUS  
 : NUMBER OF SEQUENCES: 52  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Folygy & Iardne:  
 : STREET: 1800 Diagonal Road, Suite 500  
 : CITY: Alexandria  
 : STATE: VA

```

: ZIP: 22313-0233
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patent Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/05/232,463
:

```

Query Match 2.3%; Score 32.8; DB : Length 7218;  
Best Local Similarity 3.6%; Pred. No. 4.2;  
Matches 13; Conservative 193; Mismatches 160; Indels 0

QY 1051 GGTAAGACITTTGGCTTCGAACAATGGGA:GCCTACATGCTTCTGAGTCAATGCGCAA 1110

[illegible]

RES:11.T 11

US-09-309-572-3/C  
; Sequence 3, Application US/09309572  
; Patent No. 6440730

```

: GENERAL INFORMATION:
: APPLICANT: Heinrich-Pette-Institut.
: TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
: FILE REFERENCE: P50489
: CURRENT APPLICATION NUMBER: US/09/309,572
: CURRENT FILING DATE: 1998-05-11
: EARLIER APPLICATION NUMBER: DE 198 56 463
: EARLIER FILING DATE: 1998-11-26
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 3
: LENGTH: 3376
: TYPE: DNA
: ORGANISM: Lymphocytic choriomeningitis virus
: NCBI: 309-572-3

```

Query Match	2.28	Score 32.4	DB 4	Length 3376
Best Local Similarity	54.1%	Pred. No. 3.5		
Matches	56	Conservative	0	Mismatches 56
				Indels
				Gaps

QY	426	AAATGGCCCTTCATCAACCCCAAAACGGACCGATCATGTC	CAATGGCGCGAGAAAGGTG	485
Db	2:41	AANTAGCAATTTC	CAACACAGACGGG	2082
QY	486	ATGTGTCGGTGTCTATA:CGAATCCATGTGCCCCGGCGGTGTGATCCCTACGGCACTG		545
Db	2081	ATCAAAACAACATTC	ACGCAAGATTC	2022
QY	546	GC	547	
Db	2021	TC	2020	

## RESULT 12

US-08-056-200-93  
; Sequence 93, Application US/08056200  
; Patent No. 5616500

APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-056-200-93

Query Match 2.2% Score 32.4; DB 1; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 3 CGGAACTCCATGTGGCGGTGATCTGTCAGCAGGATATTCGATGATCCAGCGGCC 62  
DB 3064 CGAGAGCCCAAGACAGAGTGTCCAGGAGGAAGAAGAAAGAGTGGAGGAGCGGAG 3123  
QY 63 GCACAGCGCTGTGGGTAAATGATAAGGCGCTGTTGTAGAAAGCGTGACCCAAACAG 122  
DB 3124 ACAGTGTCTCGGAAGGAAGAAGATTGCAGGAAGAGAGCGCGGCGCAAGAGAG 3183  
QY 123 CTCTCTGATGATCTTTTAAATGCGTGTCTATCTGGCTCTGTAA 164  
DB 3184 CTCAGGAGGAAGAGACAGCACTACGGAAGCTGGAGCGCAA 3225

RESULT 13  
US-08-800-644-93  
Sequence 93, Application US/08800644  
Patent No. 5958752  
GENERAL INFORMATION:  
APPLICANT: Steiert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/056,200  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-800-644-93

Query Match 2.2% Score 32.4; DB 2; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 3 CGGAACTCCATGTGGCGGTGATCTGTCAGCAGGATATTCGATGATCCAGCGGCC 62  
DB 3064 CGAGAGCCCAAGACAGAGTGTTCAGGAGGAAGAAGAAAGAGTGGAGGAGCGGAG 3123  
QY 63 GCACAGCGCTGTGGGTAAATGATAAGGCGCTGTTGTAGAAAGCGTGACCCAAACAG 122  
DB 3124 ACAGTGTCTCGGAAGGAAGAAGATTGCAGGAAGAGAGCGCGGCGCAAGAGAG 3183  
QY 123 CTCTCTGATGATCTTTTAAATGCGTGTCTATCTGGCTCTGTAA 164  
DB 3184 CTCAGGAGGAAGAGACAGCACTACGGAAGCTGGAGCGCAA 3225

RESULT 14  
US-08-374-686-4/C  
Sequence 4, Application US/08374686  
Patent No. 5616474  
GENERAL INFORMATION:  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu



GenCore version 5.1.6  
Copyright (c) 1995 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 06:07:31 : Search time 273 Seconds

(without alignments)  
3245237 Million cell updates/sec

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtggc.....aatgcaattatttgaacc 1442

Scoring table: IDENTITY-NUC

Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq-length: C

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*

1: /cqn2\_6/ptodata/2/pubpna/USC7\_PUBCOMB.seq.\*  
2: /cqn2\_6/ptodata/2/pubpna/PCI\_NEW\_PUB.seq.\*  
3: /cqn2\_6/ptodata/2/pubpna/US56\_NEW\_PUB.seq.\*  
4: /cqn2\_6/ptodata/2/pubpna/US56\_PUBCOMB.seq.\*  
5: /cqn2\_6/ptodata/2/pubpna/US57\_NEW\_PUB.seq.\*  
6: /cqn2\_6/ptodata/2/pubpna/PCIUS\_PUBCOMB.seq.\*  
7: /cqn2\_6/ptodata/2/pubpna/US98\_NEW\_PUB.seq.\*  
8: /cqn2\_6/ptodata/2/pubpna/US98\_PUBCOMB.seq.\*  
9: /cqn2\_6/ptodata/2/pubpna/US99\_NEW\_PUB.seq.\*  
10: /cqn2\_6/ptodata/2/pubpna/US99\_PUBCOMB.seq.\*  
11: /cqn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cqn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cqn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cqn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1442	100.0	1442	9	US-10-086-082-1
C 2	37.6	2.6	3902	9	Sequence 1, Appli
C 3	36.2	2.5	461	10	Sequence 88, Appl
C 4	35.8	2.5	941	9	Sequence 411, App
C 5	35.2	2.4	429	9	Sequence 484, Appl
C 6	35	2.4	724	9	Sequence 60, Appl
C 7	34.4	2.4	596	9	Sequence 310, App
C 8	34.4	2.4	596	9	Sequence 310, App
C 9	34.2	2.4	3188	9	Sequence 14, Appl
C 10	34	2.4	4153	9	Sequence 123, App
C 11	33.8	2.3	481	10	Sequence 372, App
C 12	33.8	2.3	1068	9	Sequence 3151, Ap
C 13	33.8	2.3	1839	10	Sequence 6010, Ap
C 14	33.8	2.3	3309400	9	Sequence 1, Appli
C 15	33.6	2.3	284	10	Sequence 2052, Ap
C 16	33.2	2.3	415	10	Sequence 2327, Ap
C 17	33.2	2.3	671	9	Sequence 346, App
C 18	33.2	2.3	671	9	Sequence 346, App
C 19	33.2	2.3	1049	9	Sequence 358, App

20 33.2 2.3 1521 9 US-10-166-037-5 Sequence 5, Appli  
21 33.2 2.3 3201 9 US-09-423-126-9 Sequence 9, Appli  
22 33.2 2.3 3201 10 US-09-280-197-7 Sequence 7, Appli  
23 33 2.3 320 10 US-09-864-761-29598 Sequence 29598, A  
24 33 2.3 521 10 US-09-864-761-13036 Sequence 13036, A  
25 32.8 2.3 383 9 US-10-040-739-1257 Sequence 1257, Ap  
26 32.8 2.3 951 10 US-09-815-242-7836 Sequence 7836, Ap  
27 32.8 2.3 5907 9 US-09-835-9768-11 Sequence 11, Appl  
28 32.5 2.3 2616 9 US-10-227-255A-2 Sequence 2, Appli  
29 32.5 2.3 2616 9 US-09-812-143-7 Sequence 7, Appli  
30 32.5 2.3 2616 9 US-09-812-143-8 Sequence 8, Appli  
31 32.5 2.3 3324 9 US-09-812-143-5 Sequence 5, Appli  
32 32.6 2.3 3324 9 US-09-812-143-6 Sequence 6, Appli  
33 32.4 2.2 4527 10 US-09-901-940-3 Sequence 3, Appli  
34 32.4 2.2 10417 9 US-10-152-040-28 Sequence 28, Appl  
35 32 2.2 1189 9 US-09-925-299-135 Sequence 135, App  
36 32 2.2 1189 10 US-09-925-299-135 Sequence 135, App  
37 32 2.2 1602 9 US-10-205-522-117 Sequence 117, App  
38 31.8 2.2 1239 9 US-10-156-761-6205 Sequence 6205, Ap  
39 31.8 2.2 9025608 9 US-10-156-761-1 Sequence 1, Appli  
40 31.5 2.2 500 9 US-09-918-995-22134 Sequence 22134, A  
41 31.5 2.2 524 9 US-10-184-644-204 Sequence 204, App  
42 31.5 2.2 524 9 US-10-184-634-204 Sequence 204, App  
43 31.5 2.2 599 9 US-10-060-036-43 Sequence 43, Appl  
44 31.5 2.2 847 9 US-10-153-668-283 Sequence 283, App  
45 31.5 2.2 957 10 US-09-974-300-983 Sequence 983, App

#### ALIGNMENTS

RESULT 1  
US-10-086-082-1  
Sequence 1, Application US/13086082  
Publication No. US2003008742A1  
GENERAL INFORMATION:  
APPLICANT: Briceon, Walter  
APPLICANT: Naughton, Andrew  
APPLICANT: Robins, Karen  
APPLICANT: Shaw, Nicholas  
APPLICANT: Zimmermann, Thomas  
TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)  
TITLE OF INVENTION: 3,3,3-TRIFLUORO-2-HYDROXY-2-METHYLPHOSPHONIC ACID  
FILE REFERENCE: 32213  
CURRENT APPLICATION NUMBER: US/10/086,082  
CURRENT FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US/05/214,675  
PRIOR FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version: 3.0  
SEQ ID NO 1  
LENGTH: 1442  
TYPE: DNA  
ORGANISM: Klebsiella oxytoca  
US-10-086-082-1

Query Match 100.0%; Score 1442; DB 9; Length 1442;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 CCGGGAACCTCCATCGCGCGTCATCTCGTCCAGCAGGATATTCGATGATCCAGCGG 60  
Db 1 CCGGGAACCTCCATCGCGCGTCATCTCGTCCAGCAGGATATTCGATGATCCAGCGG 60  
Oy 61 CCGGGAACCTCCATCGCGCGTCATCTCGTCCAGCAGGATATTCGATGATCCAGCGG 120  
Db 61 CCGGGAACCTCCATCGCGCGTCATCTCGTCCAGCAGGATATTCGATGATCCAGCGG 120  
Oy 121 AGCTCTGATGATCTTTTAAATCGCGTCGTCATCTGCTCTGTAACCTAAACGCTATAAAT 180  
Db 121 AGCTCTGATGATCTTTTAAATCGCGTCGTCATCTGCTCTGTAACCTAAACGCTATAAAT 180

QY 181 ACGTGGAGATACATATGAATGGTTGGAGAAATCCATTATGCCCCAAACGGCGTGTTCG 240  
DB 181 ACGTGGAGATACATATGAATGGTTGGAGAAATCCATTATGCCCCAAACGGCGTGTTCG 240  
QY 241 TGGCGGCGTAAACCGGTAACGATACCTGAGCGAAGAAATGCAAAAGAGTTTCATTA 300  
DB 241 TGGCGGCGTAAACCGGTAACGATACCTGAGCGAAGAAATGCAAAAGAGTTTCATTA 300  
QY 301 CACCATGGCGCTTATTCACACCGCTGCTGACATCGAACCGGTGACGGGATATTGT 360  
DB 301 CACCATGGCGCTTATTCACACCGCTGCTGACATCGAACCGGTGACGGGATATTGT 360  
QY 361 CGACACTCGAGATGCTTTTGAAGTGCTATCAATTGGAACACAGAAATTTCGAGAGAGTT 420  
DB 361 CGACACTCGAGATGCTTTTGAAGTGCTATCAATTGGAACACAGAAATTTCGAGAGAGTT 420  
QY 421 GCTAAANATGCGCTTCTCAACCCNCAAAACGACCGATCATGGTCANTGCGCGGAGAA 480  
DB 421 GCTAAANATGCGCTTCTCAACCCNCAAAACGACCGATCATGGTCANTGCGCGGAGAA 480  
QY 481 AGGTGATGTGTGCTGTCTATATCGAATCCATGTGCGCGGCGGTTCGATGCTTAGG 540  
DB 481 AGGTGATGTGTGCTGTCTATATCGAATCCATGTGCGCGGCGGTTCGATGCTTAGG 540  
QY 541 CATCTGCGCATGATCCGCAATTTTGGCGACTGACCGGACCGACGACGTCGCTGCT 600  
DB 541 CATCTGCGCATGATCCGCAATTTTGGCGACTGACCGGACCGACGACGTCGCTGCT 600  
QY 601 CAATGATCCGCTGCCCAAAAGGTGGCATGATTAACCTCGACAGTCAAAAGTCTACTG 660  
DB 601 CAATGATCCGCTGCCCAAAAGGTGGCATGATTAACCTCGACAGTCAAAAGTCTACTG 660  
QY 661 GAGCAAAACGCAATACGCTTCCCTATAAACCCTATTTGGCACCTTGAGCGCTATCGGAGA 720  
DB 661 GAGCAAAACGCAATACGCTTCCCTATAAACCCTATTTGGCACCTTGAGCGCTATCGGAGA 720  
QY 721 AATTGACTCAATCAATTCATGACGCCAGACAATACAGCGCGGAGAAATATGGAATGCGGGA 780  
DB 721 AATTGACTCAATCAATTCATGACGCCAGACAATACAGCGCGGAGAAATATGGAATGCGGGA 780  
QY 781 TATAGGACGAGGAGTATTACCTATCTGCCGTACGTGCGCTGGAGCGCGCTGTATTAT 840  
DB 781 TATAGGACGAGGAGTATTACCTATCTGCCGTACGTGCGCTGGAGCGCGCTGTATTAT 840  
QY 841 TGGTGTATGCCCATGCTTGTGAGGTGATGGTGAGATTTGCGGACCCGACGTAGAGTTTCG 900  
DB 841 TGGTGTATGCCCATGCTTGTGAGGTGATGGTGAGATTTGCGGACCCGACGTAGAGTTTCG 900  
QY 901 CTCATCACCACCATCAAAATGCAATTTGATCAAGAACTGGCACTTCTGCGGACCAAT 960  
DB 901 CTCATCACCACCATCAAAATGCAATTTGATCAAGAACTGGCACTTCTGCGGACCAAT 960  
QY 961 GGAAATGCGCGAANAATATGAGTATTGGCAGTGGAGTGGCTGGAGGATGCGAGCGG 1020  
DB 961 GGAAATGCGCGAANAATATGAGTATTGGCAGTGGAGTGGCTGGAGGATGCGAGCGG 1020  
QY 1021 AATTGCATATCGGACATTAATTTACTGGCTGGTAGAAGACTTTGGCTTCAACAATGGGA 1080  
DB 1021 AATTGCATATCGGACATTAATTTACTGGCTGGTAGAAGACTTTGGCTTCAACAATGGGA 1080  
QY 1081 TGCCTACATGCTTCTGAGTCAATGGCGCAAGTGGCTGGGCAACATGCTGCAACCGCAA 1140  
DB 1081 TGCCTACATGCTTCTGAGTCAATGGCGCAAGTGGCTGGGCAACATGCTGCAACCGCAA 1140  
QY 1141 ATACACCGTGGCGCGATGTGACAAACAAACCTTTTAACTTAGTAGAAATAACTAAACGG 1200  
DB 1141 ATACACCGTGGCGCGATGTGACAAACAAACCTTTTAACTTAGTAGAAATAACTAAACGG 1200  
QY 1201 TGAACATTACCGCGGATGTATATCGGCTTAATGTGTAAGTCAAAACATCTCTATTITAA 1260  
DB 1201 TGAACATTACCGCGGATGTATATCGGCTTAATGTGTAAGTCAAAACATCTCTATTITAA 1260  
QY 1261 CAGCTAAGCANGTGTGATATGGGGCCAGNATACACCCCAACATATTTGTTTACTTACTCC 1320

DB 1261 CAGCTAAGCANGTGTGATATGGGGCCAGNATACACCCCAACATATTTGTTTACTTACTCC 1320  
QY 1321 TTCAGGGAGTACGCGGCGCACAAAGAGTTGTCACAAATGCGGGAGCAACCCAGGCTATT 1380  
DB 1321 TTCAGGGAGTACGCGGCGCACAAAGAGTTGTCACAAATGCGGGAGCAACCCAGGCTATT 1380  
QY 1381 GCGGAAATTAATCAAAATGCGGCGCATCAACGGCAGACCACTCAATGCAATTCATTGGAT 1440  
DB 1381 GCGGAAATTAATCAAAATGCGGCGCATCAACGGCAGACCACTCAATGCAATTCATTGGAT 1440  
QY 1441 CC 1442  
DB 1441 CC 1442

RESULT 2  
US-10-108-605-88/c  
Sequence 88, Application US/10108605  
Patent No. US20020160934A1  
GENERAL INFORMATION:  
APPLICANT: Broadus, Julie  
APPLICANT: Stam, Lynn  
APPLICANT: Bachmann, Jane  
APPLICANT: Kamdar, Kim  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF  
FILE REFERENCE: 3113B  
CURRENT APPLICATION NUMBER: US/10/108,605  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: US 09/763,142  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/176,418  
PRIOR FILING DATE: 2000-01-14  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 88  
LENGTH: 3902  
TYPE: DNA  
ORGANISM: Drosophila melanogaster  
US-10-108-605-88

Query Match 2.6%; Score 37.6; DB 9; Length 3902;  
Best Local Similarity 48.6%; Pred. No. 0.33;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 320 ACACCGCTGCTGACCAATCGAACCCGCTGACCGGATATTTGCGACACTCGAGATGCTTTT 379  
DB 1458 ACACCGGACTGCTGCTGCGTATGCGAGACTGGAATTTGCTGCTCGAGTGGCTATT 1399  
QY 380 GAAGGTGCTATCAATTCGGAACAGGATATTCGGAGCCAGTTGCTAAAAATGCCCTTTCTC 439  
DB 1398 GGAGCTCTTCTCTCCGCGGGGATGCTCATTTGAACCTCTGACAGCTCCCGGCC 1339  
QY 440 AACCCACAACAGGACCGGATCATGTCATGCGGGGAGAAAGGTGATGTCGCTGTC 499  
DB 1338 GGCCGTAAAGTTGTGCTAGCGATGCGCCAAACATGCTCCCAAGTGGCCTCATGATGCTG 1279  
QY 500 TATATCGAATCATGTTGCCCGCGCGCTTGA 531  
DB 1278 GTACTCGGCTCATCTCGCCGCTGAGTGCA 1247

RESULT 3  
US-0-730-444-411/c  
Sequence 411, Application US/0977044  
Patent No. US2002023283A1  
GENERAL INFORMATION:  
APPLICANT: Gorlach, Jörn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.





Query Match	2.48;	Score 35;	DB 9;	Length 724;
Best Local Similarity	8.88;	Pred. No. 0.81;		
Matches 51;	Conservative 173;	Mismatches 357;	Indels 0;	Gaps 0;
Query Match	2.48;	Score 34.4;	DB 9;	Length 596;
Best Local Similarity	18.7%;	Pred. No. 1.1;		
Matches 74;	Conservative 112;	Mismatches 207;	Indels 2;	Gaps 2;

```
QY 406 TATTCGGAGCGGTGCTTAAATATGCGCTTTTCAGACCCGCAAAACGACGATCATGGT 445
: || || || || || || || || || || || || || || || || || || || || || ||
Db 589 K.KT.SSB.YYW.WN.SWY.YD.TCDNCC.C.C.CDN.CD.DRBTAN.ANY.S.SNYBG. 530
: || || || || || || || || || || || || || || || || || || || || || ||
QY 466 CAATCGCGCGGAGAAAGTGTGCTCGCTGTCTATATCAATCAATCCAGTTCGCGCGG 525
: || || || || || || || || || || || || || || || || || || || || || ||
Db 529 .CT..CBTBBB.A.....W.B.SCC.MT.SBIATSDSAAUKA.TTACSTCSAASS 470
: || || || || || || || || || || || || || || || || || || || || || ||
QY 526 CGTGTATCCCTACGGCATCTCGCGCATGATCCGCATTTTGGGGAGCT-GACCGGGACCG 584
: || || || || || || || || || || || || || || || || || || || || || ||
Db 469 CSNATANTCSSAASS.SNABAS.CSSBASS.SNATAST.SSAASSSHSNATASTCSSAASS 410
: || || || || || || || || || || || || || || || || || || || || || ||
QY 585 ACCTGACGGCCATCTCAATGATCGCGTCCAGAAAGGTCGCGATGATTAATCAATCGACA 644
: || || || || || || || || || || || || || || || || || || || || || ||
Db 409 .SNATASBCSSAASS.SNATASTCSSBASS.SNATANTCSSAASS.SNATANTCSS.ASS 350
: || || || || || || || || || || || || || || || || || || || || || ||
QY 645 GTGAAAGGCTACTGGAGCAAGCGCAATGCGTTCCTATATAACCCCATATTTGGCAAC 703
: || || || || || || || || || || || || || || || || || || || || || ||
Db 349 .SNABAS.CSSBASS.SNATACTCSSAASSSHSNATASTCSSAASS.SNATASTCSSAASS 290
: || || || || || || || || || || || || || || || || || || || || || ||
QY 704 TTGACGGTATGCCAGAAATTAATCAATCAATCAATGACGCCAGACAAATACGGCGGG 763
: || || || || || || || || || || || || || || || || || || || || || ||
Db 289 .SNATASTCSSBASSSHSNATASTCSSAASS.SNATACTCSAAAYS.SNATASTCSSAASS 230
: || || || || || || || || || || || || || || || || || || || || || ||
QY 764 AATATGGATGTCGGGATAGAGGACCGAGGATAT 798
: || || || || || || || || || || || || || || || || || || || || || ||
Db 229 .SNATASTYSSBASS.SNATASTCSSAASS.SNAT 195
: || || || || || || || || || || || || || || || || || || || || || ||
```

## RESULT 6

US-10-184-634-310/C

Sequence 310, Application US/10184634

Publication No. US20030068694A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430RIC217

CURRENT APPLICATION NUMBER: US/10/184,634

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 310

LENGTH: 596

TYPE: PRT

ORGANISM: Homo Sapien

US-10-184-634-310

Query Match 2.4%; Score 34.4; DB 9; Length 596;

Best Local Similarity 18.7%; Pred. No. 1.1;

Matches 74; Conservative 112; Mismatches 207; Indels 2; Gaps 2;

```
QY 406 TATTCGGAGCGGTGCTTAAATATGCGCTTTTCAGACCCGCAAAACGACGATCATGGT 465
: || || || || || || || || || || || || || || || || || || || || || ||
Db 589 K.KT.SSB.YYW.WN.SWY.YD.TCDNCC.C.C.CDN.CD.DRBTAN.ANY.S.SNYBG. 530
: || || || || || || || || || || || || || || || || || || || || || ||
QY 466 CAATCGCGCGGAGAAAGTGTGCTCGCTGTCTATATCAATCAATCCAGTTCGCGCGG 525
: || || || || || || || || || || || || || || || || || || || || || ||
Db 529 .CT..CBTBBB.A.....W.B.SCC.MT.SBIATSDSAAUKA.TTACSTCSAASS 470
: || || || || || || || || || || || || || || || || || || || || || ||
QY 526 CGTGTATCCCTACGGCATCTCGCGCATGATCCGCATTTTGGGGAGCT-GACCGGGACCG 584
: || || || || || || || || || || || || || || || || || || || || || ||
Db 469 CSNATANTCSSAASS.SNABAS.CSSBASS.SNATASTCSSAASS.SNATASTCSSAASS 410
: || || || || || || || || || || || || || || || || || || || || || ||
```

## RESULT 9

US-10-027-828-14/C

Sequence 14, Application US/0027828

Publication No. US20030013656A1

GENERAL INFORMATION:

APPLICANT: Liedtke, Wolfgang

APPLICANT: Heller, Stefan

APPLICANT: Hudsabeth, Albert J.

APPLICANT: Friedman, Jeffrey M.

TITLE OF INVENTION: VR-OAC, AN OSMOTICALLY ACTIVATED CHANNEL PROTEIN, NUCLEIC ACI

FILE REFERENCE: IT, AND USES THEREOF

CURRENT APPLICATION NUMBER: US/10/027,828

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 20

SOFTWARE: PatentIn version 3.1

SEQ ID NO 14

LENGTH: 3188

TYPE: DNA

ORGANISM: Mus musculus

US-10-027-828-14

Query Match 2.4%; Score 34.2; DB 9; Length 3188;

Best Local Similarity 51.7%; Pred. No. 4;

Matches 78; Conservative 0; Mismatches 73; Indels 2; Gaps 3;

```
QY 2 CGGGAAC:CCATGTGGCGGTGATCTGCTGCGAGCAGATATTCGATGATCGAGCGGC 61
: || || || || || || || || || || || || || || || || || || || || || ||
Db 2025 CGTGCAGTGTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 1970
: || || || || || || || || || || || || || || || || || || || || || ||
QY 62 CGCACAGCTCTGTGGGTAATGGAATGAGCCCTGGTTGTAGAACGCTGACCAACACA 121
: || || || || || || || || || || || || || || || || || || || || || ||
Db 1969 TGACCGAGGCTGAGGCA:ACCCGATCATGAAGCAGAGGTATACACGACGAGGAAGA 1910
: || || || || || || || || || || || || || || || || || || || || || ||
QY 122 GCTCTCTGATGATCTTTTAAATGCGCTGCTCAT 152
: || || || || || || || || || || || || || || || || || || || || || ||
Db :999 GGTCTTTGAAGAGGATCTTCTGATCATGAT 1879
: || || || || || || || || || || || || || || || || || || || || || ||
```

## RESULT 10

US-10-153-668-123

Sequence 123, Application US/10153668

Publication No. US20030092616A1

GENERAL INFORMATION:

APPLICANT: HONDA, Goichi

APPLICANT: MATSUDA, Akio

APPLICANT: MURAMATSU, Shuji

APPLICANT: ISHIZAWA, Kenya

TITLE OF INVENTION: STAT6 Activating Gene

FILE REFERENCE: 1254-0207P

CURRENT APPLICATION NUMBER: US/10/153,668

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 24

SOFTWARE: PatentIn version 3.1

SEQ ID NO 14

LENGTH: 3188

TYPE: DNA

ORGANISM: Mus musculus

US-10-027-828-14

```

: PRIOR FILING DATE: 2001-05-25
: PRIOR APPLICATION NUMBER: US 60/316,031
: PRIOR FILING DATE: 2001-08-31
: PRIOR APPLICATION NUMBER: US 60/428,403
: PRIOR FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: JP 2001-157643
: PRIOR FILING DATE: 2001-05-25
: PRIOR APPLICATION NUMBER: JP 2001-260681
: PRIOR FILING DATE: 2001-08-30
: PRIOR APPLICATION NUMBER: JP 2001-31375
: PRIOR FILING DATE: 2001-10-10
: NUMBER OF SEQ ID NOS: 488
: SOFTWARE: SeqMan Ver. 2.0
: SEQ ID NO: 123
: LENGTH: 4153
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (175)..(3024)
: US-10-153-668-123

```

	Query Match	2.4%	Score 34:	DB 9:	Length 4153:
	Best Local Similarity	52.9%:	Pred. No. 5.6:		
Matches	73:	Conservative	0:	Mismatches	65:
				Indels	0:
				Gaps	0:
QY	196	TATGAANTGGTGGAGAAATCCATTATGGCCAACGCGGTCTTGTTGCCCGGGTAAAC	255		
Dd	1074	TATTAANTGTTTGATGCACCATTATGCTTCGCGGTTCCTAAGNATAGCCTTCACA	1133		
QY	256	GGTAAGCCATCACTGACGAGAGAATGCAANAAGAGTTTCA-TATACCATATGGCGCTTA	315		
Dd	1134	GTCAGAGAGGACTGCTGGAGCAATAGAGANCAITCTGCTTACAGCACTCACTACTG	1193		
QY	316	TTCCACACCCGTCCTGCAC	333		
Dd	1194	TTCCGAGGACCACTGCAC	1211		

```

RESULT 11
US-09-833-381-372/c
; Sequence 372, Application US/09833381
; Patent No. US20020132030A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 372
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: {1}...{481}
; OTHER INFORMATION: n A,T,C or G
US-09-833-381-372

```

	Query Match	2.3%	Score 33.8:	DB 10:	Length 481:
	Best Local Similarity	51.7%:	Pred. 0.1.6:		
	Matches 77:	Conservative	0:	Mismatches 72:	Indels 0:
QY	16	TGGCGGATGCTTGTGCAGACAGGATATTGGATGATCCAGCGGAGCGCACAGGGCCTGG	75		
Dd	172	TGGCGGTACTTTATCACITCAGCATATGGCGGTSTTGTCAGAGATTGGCGATGGTAGC	115		
QY	76	CGGTATGGATAAGGCCTGGTGTGTAGAAGCGGTGAGCGTAACACACAGCTCTCGATGATC	135		

```

Db      112  TGTGATGATCATAGCGGAGCGGTGGAAACGGGTACCTCGAACACAGATTTTTCATGCAC
Qy      136  TTTTAATGCTGCTGCTATCTGGCTCTGTAA 164
          | | | | | | | | | | | | | | | |
Db      52  CGCAACAICCTTACACCCCGTGCGCTGTIA 24

RESULT 12
US-09-738-626-3151
: Sequence 3151, Application: US/09738626
: Publication No. US20020157605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIALI, KEIKO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738,626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 3151
: LENGTH: 1068
: TYPE: DNA
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-3151

```

Query Match:	2.3%	Score 33.8;	DB 9;	Length 1058;
Best Local Similarity	64.9%;	pred. No. 2.7;		
Matches	50;	Conservative 0;	Mismatches 27;	Indels 0; Gaps 0
Qy	81:	GGTACGTGCGCGCTGGAGGCCCGCTGTTATTGTGTGATGCCCATCGCTTCAGGGGTGA	GG	870
Db	392	GGTCCCTGTGCGTGGGAAGTGGCGTTTACCTGTGTGATGCCCGCTTCGCGTGGGTGG	GG	451
Qy	871	TCGATGTTGCGGGACCG	887	
Db	452	ATAGGTTTGGTGGCCG	468	

```

RESULT 13
: 05-09-815-242-601C
: Sequence #G10, Application US/09815242
: Patent No. US2002061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCES: ELITRA.G11A
: CURRENT APPLICATION NUMBER: US/09/815.242
: CURRENT FILING DATE: 2302-03-21
: PRIOR APPLICATION NUMBER: 69/191,078
: PRIOR FILING DATE: 2000-03-21

```

PRIOR APPLICATION NUMBER: 60/255,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/237,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,952  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6010  
LENGTH: 1839  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1839)  
US-09-815-242-6010

Query Match 2.3% Score 33.8; DB 10; Length 1839;  
Best Local Similarity 51.7%; Pred. No. 3.8;  
Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 16 TGGCGGTATCTTATACCTCAGGATATTCGATGATCCAGCGCGGCACACGCGTGTG 75  
DB 626 TGGCGGTATCTTATACCTCAGGATATTCGATGATCCAGCGCGGCACACGCGTGTG 685  
QY 76 CGGTAATGGATAAGGCTTGTAGAACGCTGACCCCAACACACGCTCTCTGATGAT 135  
DB 686 TGGTATGATATCAGGCGGAGCGGTGGAAACGGGTACCGTGCACACAGATCTTTCATGCA 745  
QY 136 TTTAATGGTGTCTATCTGTCTGTCTGTAA 164  
DB 746 CGCAACATCTTACACCGGTGCGCTGTATA 774

RESULT 14  
US-09-738-626-1/C  
Sequence 1. Application US/9738626  
Publication No. US200201976C5A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 1  
LENGTH: 3309400  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 2.3% Score 33.8; DB 9; Length 3309400;  
Best Local Similarity 64.9%; Pred. No. 3.3e+02;  
Matches 50; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 811 GGTACGTGGCGCTGGAGCGCGCTGTTTATTTGATGATGCCCATGCTTGTCAAGGTGAIGG 870  
DB 3046738 GGTCCCTGTGGCGTGGAAAGTGGCGGTTTACCTGGTGAISCGGCTTCTCGGCGTGGTGTGG 3046739  
QY 871 TCAGATTTCCGGGACCG 887  
DB 3046738 ATAGGTTTGGTCGCGCG 3046722

RESULT 15  
US-09-983-965-2052  
Sequence 2052, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Ningbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 37-21(10297)C  
CURRENT APPLICATION NUMBER: US/09/983,965  
CURRENT FILING DATE: 2001-10-26  
PRIOR APPLICATION NUMBER: US 09/465,231  
PRIOR FILING DATE: 1999-12-15  
PRIOR APPLICATION NUMBER: US 60/113,678  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 5912  
SEQ ID NO 2052  
LENGTH: 284  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
OTHER INFORMATION: Clone ID: 28-LIB1057-005-Q1-K1-G7  
US-09-983-965-2052

Query Match 2.3% Score 33.6; DB 10; Length 284;  
Best Local Similarity 48.9%; Pred. No. 1.3;  
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
QY 29 GTTCGAGCAGGATATTCGATGATCCAGCGCGGCACACGCGCTGTGCGGTATGATAA 86  
DB 4 GTTCGAGGCGTGTCTTCTCTGATCCGCAACCGCGCGCTGTGCGGTATGATAACT 90  
QY 89 AGSCCTGGTGTAGAAAGCTGACCCCAACACAGCTCTGTGATGATCTTTTATCGGTG 148  
DB 92 ACTCTCTCTTTGTAAAGTGTGAACCAAGGGAAGAAACACAGGAGGTTTATCGCTGC 150  
QY 149 TCATCTGGCTCTGTAACTAAAGCTATAAATACGTGGAGAAATACATATGAATGGTTG 208  
DB 151 TCAGCTTGTAGAGACAGAAAAGATTGTAATCTTTTCAATGGGAAGATGAAAAGTTGCA 210  
QY 209 GAAG 212  
DB 211 GGAG 214  
Search completed: July 6, 2003, 07:56:52  
Job time : 286 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 04:04:05 ; Search time 1972 Seconds  
(without alignments)  
11842.744 Million cell updates/sec

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtggcc.....aatgcaattcatgtgatacc 1442

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8697743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match C%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthm:\*
  - 3: em\_estin:\*
  - 4: em\_estm:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_estc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_est3:\*
  - 12: gb\_est4:\*
  - 13: gb\_est5:\*
  - 14: gb\_est6:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pla:\*
  - 21: em\_gss\_vit:\*
  - 22: em\_gss\_fus:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_othe:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39.6	2.7	280	10 AW380638	AW380638 RC2-HT027
2	37.8	2.5	551	13 BM615374	BM615374 170006871
3	36.2	2.5	533	10 AV551507	AV551507 AV551507
4	36.2	2.5	678	13 BM582997	BM582997 170006872
5	36.2	2.5	1101	17 CNS00370	AL064455 Drosophila
6	36	2.5	495	10 AV629284	AV629284 AV629284

C	7	35.6	2.5	482	17	BH767373	BH767373 BMBAC348F
C	8	35.6	2.5	735	17	AG117547	AG117547 Pan trogl
C	9	35.2	2.4	611	10	BE442656	BE442656 WHE1101.D
C	10	35.2	2.4	772	17	BH709682	BH709682 BOMEL07TF
C	11	35	2.4	266	9	AA020733	AA020733 zeb3810.S
C	12	35	2.4	421	14	H75593	H75593 yu05f04.f1
C	13	35	2.4	441	9	AA019639	AA019639 zeb2d01.S
C	14	35	2.4	492	12	AA001557	AA001557 zeb4606.S
C	15	35	2.4	505	12	BG834211	BG834211 352349.MA
C	16	35	2.4	691	17	B57961	B57961 CIT-HSP-201
C	17	34.8	2.4	328	9	AU109318	AU109318 AUI05318
C	18	34.8	2.4	356	17	AQ610392	AQ610392 HS_5101.B
C	19	34.8	2.4	422	17	AQ555596	AQ555596 HS_5229.A
C	20	34.8	2.4	475	9	AI743670	AI743670 wq41606.X
C	21	34.8	2.4	477	9	AI743687	AI743687 qp53e12.X
C	22	34.8	2.4	482	24	BQ247372	BQ247372 TAE25047G
C	23	34.8	2.4	502	9	AI289776	AI289776 qw12a01.X
C	24	34.8	2.4	524	12	BG052847	BG052847 RHI22.1A
C	25	34.6	2.4	511	17	AQ482699	AQ482699 RPI1-11-2
C	26	34.6	2.4	552	17	AQ478968	AQ478968 RPI1-11-2
C	27	34.4	2.4	410	12	BF484542	BF484542 WHE2324.G
C	28	34.4	2.4	474	10	BE496530	BE496530 WHE0971.G
C	29	34.4	2.4	483	9	AA833827	AA833827 od62c06.S
C	30	34.4	2.4	491	14	H08687	H08687 Y193e12.cl
C	31	34.4	2.4	623	17	B69197	B69197 CIT-HSP-205
C	32	34.4	2.4	641	13	BK352441	BK352441 50071-2-1
C	33	34.4	2.4	641	13	BK393654	BK393654 50072-2-1
C	34	34.4	2.4	1100	17	CNS00DYG	AL077401 Drosophila
C	35	34.2	2.4	466	13	BM136789	BM136789 WHE2617.E
C	36	34.2	2.4	531	13	BJ246954	BJ246954 BJ246954
C	37	34.2	2.4	656	13	BJ234037	BJ234037 RJ234037
C	38	34.2	2.4	681	10	BE489300	BE489300 WHE0363.B
C	39	34.2	2.4	903	13	BM018322	BM018322 603645981
C	40	34	2.4	347	14	R63461	R63461 Y108a12.S1
C	41	34	2.4	515	12	BF252201	BF252201 EST419463
C	42	34	2.4	540	12	BF252018	BF252018 EST419280
C	43	34	2.4	565	17	AQ430850	AQ430850 HS_5103.A
C	44	34	2.4	669	12	BF251067	BF251067 EST418324
C	45	34	2.4	760	12	BG574169	BG574169 602596348

## ALIGNMENTS

RESULT 1  
AW380638 280 bp mRNA linear EST 04-FSB-200C  
RC2-HT0275-011199-011-f11 HT0275 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AW380638  
VERSION  
KEYWORDS EST.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 280)  
RCGP <http://www.udwig.org.br/RCGP-ESTES>.  
The FAPESP/LICR Human Cancer Genome Project  
Unpublished (1999)  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC2&t2=RC2-HT0275-011199-011-f11&at3=1999-11-01&at4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 37

BASE COUNT:

```

RESULT 4
BM582997
LOCUS      17900667277553 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
DEFINITION 1960C449689205 5', mRNA sequence.
ACCESSION  BM582997
VERSION     BM582997.1
KEYWORDS    EST.
SOURCE      African malaria mosquito.
ORGANISM    Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE   1 (bases 1 to 678)
AUTHORS     Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE       Cella Anopheles gambiae EST project
JOURNAL     Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 24045334580
            Email: Holt.RA@celera.com
            Plate: NU01004AYC row: G column: 15
            Seq primer: M13 Reverse.
FEATURES    Location/Qualifiers
             1..678
             /organism="Anopheles gambiae"
             /strain="RSP-ST (Reduced susc. to Permethrin - std.
             chromosome)"
             /db_xref="taxon:7165"
             /clone="1960C449689205"
             /clone_lib="A.Gam.ad.cDNA.blood1"
             /dev_stage="Adult"
             /lab_host="DH10b"
             /note="Vector: pSPori1; Site 1: SalI; Site 2: NotI; whole
             adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
             hours after human blood feeding. cDNA inserts >500 bp
             cloned directionally into pSPori 1. Not 2 site is 3'.
             Clones available through the Malaria Research and
             Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT  177 a 178 c 228 g 95 t
ORIGIN
Query Match      2.5%; Score 36.2; DB 13; Length 678;
Best Local Similarity 57.5%; Pred. No. 7.7;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Cy 10 TCCATGTGGCGTGTATCTCTCCAGCAGGATATGGATGATGACGGGCGCCAGAGG 69
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 263 TCCAGCGGACGACATAGTGTGTGTAAGGACGGTGTGATGTCAGACGGGTCGACG 324
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Cy 70 GCTGTGGCGTAAAGAGCGTGTGTGAGAAAGCGTGACCCACACAG 122
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 323 ACAATCTGCTCCAGCAGCGCGGTGTATGCGGAGATGTGACACGACGAGTG 375

Query Match      2.5%; Score 36.2; DB 13; Length 678;
Best Local Similarity 57.5%; Pred. No. 7.7;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

RESULT 5
CNS00370
LOCUS      1101 bp DNA linear GSS C3-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence IET3 end of PAC #
            BACR08K14 of RP11-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL064465
VERSION     AL064465.1
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
ORGANISM    Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

```

```

REFERENCE
AUTHORS     1 (bases 1 to 1101)
GENOSCOPE   Direct Submission
TITLE       Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org/TheBDGP/Drosophila
            melanogasterBAClibrary.html
            Aaron Mammeter in Piter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RP11-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES    Location/Qualifiers
             1..1101
             /organism="Drosophila melanogaster"
             /db_xref="taxon:7227"
             /clone="BACR08K14"
             /clone_lib="RP11-98"
             /note="end : TET3"
BASE COUNT    374 a 61 c 129 t 459 others
ORIGIN
Query Match      2.5%; Score 36.2; DB 17; Length 1101;
Best Local Similarity 12.2%; Pred. No. 11;
Matches 49; Conservative 173; Mismatches 191; Indels 0; Gaps 0;

Cy 186 GAGATAACATATGAATGTTGGAGATCCATATATGGCAACCGGTTGGTGGCCG 245
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 693 KGRVWBDNVKAAITDTMTXIRAVAMARAGAAAMARTAMRAAAKAAKANKKGTG 752
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Cy 246 GGGTAACCGGTAACCATCCTGACGGAAGAAATGCAAAAGAGTTTCATTACCA 305
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 753 ATKDKMKMKMGKHKKKDKMDGAKATAMATRAGAVGAKARRAKRWAGAARRVA 812
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Cy 306 TTAGCCCTTATCCACACCGGCTCTGACCATCGAACCCTGACCGGATATTGICGAC 365
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 KGANRVKADAMRRHRHAKWANGAGHTVNVVBNMGNGHKKMGNTKKNVSVKT 872
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Cy 366 CTCGAGATGTTTGAAGTGTCTATCAATTCGGAACAGCAIATTCGACGAGTTGCTAA 425
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 873 VKBHKDKHTTTTNTNNMTHTTHVMBMBMGCMRMDMGMRGVGHRVGRKMGNNRM 932
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Cy 426 AATGCGCCCTTCACACCCCAACACCGGCGCATCATGTCGCGGCAAGAGTG 485
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 933 VMRVNVNVKAVAVMGHVMVMVMVMVMVMVMVMVMVMVMVMVMVMVMVMVMVM 992
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Cy 486 ATGTGTCGCTGCTATATCGAATCCATGTTGCCCGCGCGGCTTGATCCCTACGGCATCT 545
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 993 GNVCVBNMKVKNKMGKMGHNVTVMGKMBVBVNNMKVNNMKMGKNTKIMKMGHTGW 1052
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Cy 546 GCGCCATGATTCGCGATTTTGGCGGACTGACCGGACCGACCT 588
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1053 NMKGHVTRDWTGKGRGKGGKGRKGRKMGKGDGTGMGVGVK 1095

RESULT 6
AV629284/c
LOCUS      AV629284
DEFINITION AV629284 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LCL055h05_r 5', mRNA sequence.
ACCESSION  AV629284
VERSION     AV629284.1
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.

```





and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 226-0045, Japan  
 (E-mail: chimbes@gscl.riken.go.jp, URL: http://hgp.gscl.riken.go.jp/  
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PFB. This BAC end  
 was generated during the RAD process and may have higher chance of  
 clone tracking errors.

## PRIMERS

Sequencing: -21M13

## LIBRARY

Vector : pXSI45

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

.. 735

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PFB-125G18.F"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PFB Chimpanzee Male BAC Library"

188 a 186 c 150 g 198 t 3 others

## BASE COUNT

ORIGIN

Query Match 2.5%; Score 35.6; DB 17; Length 735;

Best Local Similarity 58.5%; Pred. No. 13;

Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1315 TACTCCITCAGCGGAGTGCACGAGGTTGTACACAGTTCGCGGAGCAACCCAG 1374

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 123 TCCTCCTCAGCGGAGGAATGAGCTCATTCCTATTTGTCGCGAGCAGCACACAG 182

QY 1375 GCTATTGCGGAATTAATCAAAATGCGGCATCAACGGCAGACACAC 1420

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 183 GTCTGGACCAATGAATTAATTCGCTGCCATTCGAATCAAGCAG 228

## RESULT 9

BE442656

LOCUS

WHE1101\_D03\_G052S wheat etiolated seedling root normalized cDNA

Library Trilicium aestivum cDNA clone WHE1101\_D03\_G05, mRNA

sequence.

BE442656.1 GI:9442089

EST.

Source bread wheat.

ORGANISM

Trilicium aestivum

Eukaryota; Viridiplantae; Strophophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae

; Triticeae; Triticum.

1 (bases 1 to 611)

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han

, P.S., Hsiao, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,

Paesch, C.J., Seaton, C.L., Tong, J.C. and Zhang, E.

The structure and function of the expressed portion of the wheat

genomes - Normalized root cDNA library

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105959773

Fax: 5105959818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: StrataGene SK primer.

Location/Qualifiers

1..611

/organism="Trilicium aestivum"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

## FEATURES

source

/clone="WHE1101\_D03\_G05"

/clone\_lib="wheat etiolated seedling root normalized cDNA

library"

/tissue\_type="Root"

/dev\_stage="Five day old etiolated seedling"

/lab\_host="E. coli DH10B"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid

pBluescript SK; Site\_1: EcoRI; Site\_2: XhoI; Seeds were

surface-sterilized, germinated and grown aseptically in

the dark at room temperature on filter paper with water,

mystatin and cefotaxime in covered crystallization

dishes. Roots were harvested. The tissue, total RNA, and

poly(A) RNA were prepared, a cDNA library was made in the

TJ Clone lab (Choi, Close, Fenton) at the University of

California, Riverside. The cDNA clones were in vivo

excised to give pBluescript phagemids before

normalization was carried out. The mass excision of

phagemid library and normalization were done in HT Nguyen

lab by D. Zhang at Texas Tech University. Normalization

protocol used was that of Soares. Plasmid DNA

preparations and DNA sequencing were performed in the OD

Anderson lab (all other authors)."

BASE COUNT 181 a 111 c 153 g 166 t

## ORIGIN

Query Match 2.4%; Score 35.2; DB 10; Length 611;

Best Local Similarity 51.2%; Pred. No. 15;

Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 719 GAATTCGACTCAATCAATCAGCCGAGCAATCACGCGGGAATATGATGCGCG 778

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 141 GCAATGAAGCAGCAAGAACTATCTCTGGAAGAGAGATGGGGGAATTTGACATAAG 200

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 779 GATAAGCAGCAGGAGTATTACCTATCTGCGGTACCTGCGCTGGAGCGCCGCTTT 836

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 201 ATCTAAGCAGAGGTTCCAAAGTTATCTACCAATTTGTGTGAAGGAGCAATCTGAGT 260

QY 839 ATTGGTATGCCATCTTGTTCAGGTCGATGCGATTT 878

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 261 ACTGGCAATATGCACITTTCCCAAGGTGATGGTGAAGTCT 300

## RESULT 10

BE709682/c

LOCUS

BE709682

DEFINITION

Brassica oleracea

sequence.

BE709682.1 GI:18797369

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;

Rosidae; eurosids 1; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 772)

Town, C.D., Van Aken, S., Unterbach, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSSs: BOMEL07P

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1..772

/organism="Brassica oleracea"

/strain="TOL000DH3"

## FEATURES

source

```
/db_xref="taxon:3712"
/clone="BOMEL07"
/clone_lib="BO_2_3_KB"
/Note="vector: PHOS1; Site_1: HstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 175 a 165 c 181 g 251 t
ORIGIN
Query Match 2.4%; Score 35.2; DB 17; Length 772;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 592 GCGATGCTCAATGATCGCTGCCAGAAAGGTGGCATGATTAAAGTCGACAGTGAAGA 65:
DB 636 GGACTTCCTGATAGAGCGGCTCATCAAGCAGAGTTCCTTAGAAGATCGTAAAGA 577
QY 652 GGTCTACTGGAGCAACGCCATAGCTTCCTTAAACCCCATATTGCGACCTTGAGCT 71:
DB 576 CGTCTAAGTCGAAACAGCGGCAAGACAGCGCTAGAACGTCGTCACAGCTATGAT 517
QY 712 ATCGCCGAAATTCGCTCAATTCATTCATCAAGCCAGACATCAAGCGGGGATA 767
DB 516 TTCATAGAAATGAGCCCATCAGCTGCTATGACGGTTTATCTTACTTCGGGAAA 46:
RESULT 1:
AA020733/c 264 bp mRNA linear EST 30-JAN-1997
LOCUS ze63b10.s1 Soares retina N2b4HR Homo sapiens cDNA clone
DEFINITION IMAGE:363643.3 similar to WP:K02A2.3 CEG279) BOMETANIDE-SENSITIVE
NA-K-Cl cotransporter 7, mRNA sequence.
ACCESSION AA020733.1 GI:1484516
VERSION EST
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 268)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le.N., Le.N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan.F., Thierry-Mieg,J., Trevisan,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63103
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2350 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 171.
Location/Qualifiers
1..268
/organism="Homo sapiens"
/db_xref="GDB:1280346"
/db_xref="taxon:9606"
/clone="IMAGE:363643"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue-type="retina"
/dev_stage="55 year old"
/lab_host="DH10R (ampicillin resistant)"
/Note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st
strand cDNA was primed with a Not 1 - cllqo(dT) primer {5'
TGTACCACTGAAAGTGGAGCGCGCGCTTTTITTTTITTTT 3'}.
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick K. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Falina Bonaldo.
BASE COUNT 50 a 72 c 71 g 71 t 4 others
ORIGIN
Query Match 2.4%; Score 35; DB 9; Length 268;
Best Local Similarity 50.6%; Pred. No. 9.8;
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 535 CTACGGCATCTGCGCATGATTCGGCATTTTGGCGGANTGACGGGACGACCTGACGCG 594
DB 167 CAAGGAGCTTCTTCAGCATCAAGCGGAGTGGGAGAACTTGAACCACTCCACAGTGGCGG 108
QY 595 CATGCTCAATGATTCGGCTGCCAGAAAGGTGGCATGATTAAACTCGACAGTGAAGAAGT 654
DB 107 CATGCACACTGCTGCTGGCTGAACGAGGTTCATCGTGAAGAAATCCCGGACGCGCAAGCT 48
QY 655 CTACTGGAGCAACGCCATACGCTTCCTATAAAGCCCATATTG 698
DB 47 TGTTCCTTCATCATGCTGGCGCTCCCGCAACCGCAATGCTG 4
RESULT 12
H75593/c 421 bp mRNA linear EST 01-NOV-1995
LOCUS yu05104.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:232927.5, mRNA sequence.
ACCESSION H75593.1 GI:1049521
VERSION EST
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le.N., Le.N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan.F., Thierry-Mieg,J., Trevisan,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert size: 1012
High quality sequence stops: 366
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1012 Std Error: 0.00
Seq primer: M13Rpl
High quality sequence stop: 366.
Location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="GDB:3786104"
/db_xref="taxon:9606"
/clone="IMAGE:232927"
/clone_lib="Soares fetal liver spleen INFLS"
```

```

/sex="male"
/dev_stage="20 week-post-conception fetus"
/lab_host="RH95 (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7D (Pharmacia) with a modified polylinker; Site: 1; Pac 1; Site 2; Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'-AAGTGGAGGATTAATTAAGATCTTATTATTTTITTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Eco RI sites of the modified pT73 vector and went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      89 a 104 c 107 g 113 t      8 others
ORIGIN

```

```

Query Match      2.4%; Score 35; DB 14; Length 421;
Best Local Similarity 51.6%; Pred. No. 13;
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1044 ACTGGCTGTAGACACTTGGCTTCGACACATGGGATCGCTACATGCTTCGAGTCAT 1103
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 242 ACTGGACAGGAGTACCTTGGCTGGCAGGAGATCATCTCAACCCAGCTGTGACACAT 183
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1104 GCGGCAAGTGGCGCTGGCCACATCGGTCGACCCCAATACACCTTGGCGGATGCGA 1163
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 182 CCCACAGGAGTCTTGGGCGGAATGGCGGTCGCACAGTTACACAGTTGGGGTGAACC 123
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1164 ACAAAACCTGTTAGTTAGTAGGAAATACATAACC 1198
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 122 ACTATACCTTTAATAGAGCTGATCTTCTACCC 88
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 13
AA019639/c
LOCUS      ze62q01.s1 Soares retina N2b4HR Homo sapiens cDNA clone
DEFINITION IMAGE:363600 3' similar to WP:K02A2.3 CE02791 BUMETANIDE-SENSITIVE
NA-K-Cl COTRANSPORTER 1, mRNA sequence.
ACCESSION AA019639
VERSION    AA019639.1 GI:1453067
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, P.,
            Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gier, W., Hawkins,
            M., Hultman, M., Kucaba, T., Lucy, M., Le, M., Le, N., Nardis, E., Moore,
            B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
            Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
            Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.,
            Generation and analysis of 28C.000 Human expressed sequence tags
            Genome Res. 5 (4): 837-828 (1996)

```

```

TITLE       Contact: Wilison RK
JOURNAL     Washington University School of Medicine
MEDLINE     4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT      Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2010 Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 360.
            Location/Qualifiers
            1. .441
            /organism="Homo sapiens"
            /db_xref="GDB:1280303"
            /db_xref="taxon:9606"
            /clone="IMAGE:363600"
            /clone_lib="Soares retina N2b4HR"

```

```

FEATURES
Source

```

```

/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="CH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: pT7D (Pharmacia) with a modified polylinker; Site: 1; Not 1; Site 2; Eco RI; 1st strand cDNA was primed with a Not 1 - oligo(dT) primer [5'-GTACCAATCTGAAGTGGAGGCGGCCCTTTTITTTTITTTT 3']; double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not 1 and cloned into the Not 1 and Eco RI sites of a modified pT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)- RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Frederick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      65 a 137 c 120 g 113 t      6 others
ORIGIN

```

```

Query Match      2.4%; Score 35; DB 9; Length 441;
Best Local Similarity 50.6%; Pred. No. 14;
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 535 CTACGGCATCGCCCAATGATTCGGCATTTTGGCGGACTACCGGGACCCACCTGACGGC 594
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 CAAGGACTCTTCAGCATGAAGCGGAGTGGGAGAACTTGAACCAAGTCCCAACGTGGCGC 108
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 595 CATGCTCAATGATCGCTGCCAGAAAAGTGGCGCATGATTAAACTTCACAGTGAAGAGGT 654
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 107 CATGCACTGTCGCGCTGACGAGGATCATCGTCAAGAAATCCCGGACCCCAAGCT 48
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 655 CTACTGGAGCAAGCCCATAGCTTCCCTATAAACCCCATATTC 698
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 47 TGTTCGTCACATGCTGCGCTCCCGCAACCGCAATGGT 4
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 14
AA01557/c
LOCUS      ze62d06.s1 Soares retina N2b4HR Homo sapiens cDNA clone
DEFINITION IMAGE:362027 3' similar to WP:K02A2.3 CE02791 BUMETANIDE-SFNSITIVE
NA-K-Cl COTRANSPORTER ; contains Alu repetitive element, mRNA
ACCESSION AA01557
VERSION    AA01557.1 GI:1437022
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Hillier, L., Clark, N., Dubuque, T., Eliason, K., Hawkins, M., Holman,
            M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
            Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston,
            R., Williamson, A., Wohlmann, P. and Wilson, R.,
            The WashU-Merck EST Project
            Unpublished (1995)
            Contact: Wilison RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 3433 Std Error: 0.00
            Seq primer: mob.REGA+PT
            High quality sequence stop: 233.
            Location/Qualifiers
            1. .492
            /organism="Homo sapiens"
            /db_xref="GDB:1278730"

```

```

FEATURES
Source

```

```

/db_xref="taxon:9606"
/clone="IMAGE:362027"
/clone_lib="Soares retina N2b4R"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Ordan; eye; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer; 15'
TGTTACCAATCIGAGTGAGCGCGCTTTTITTTTITTTT 3';
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)- RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Patima Bonaldo."
BASE COUNT      80 a 141 c 131 g 125 t      1; others
ORIGIN
Query Match      2.4%; Score 35; DB 9; Length 492;
Best Local Similarity 50.6%; Pred. No. 15;
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 535 CTACGGCATCTGGCCATCTTCGCATTTTGGGAGTGTGACGGAGCGAGCGGC 594
DB 167 CAAGACTTCTTCCACATGAACCGGAGTGGGAACTTGACACGATCCACGTTGGCGG 100
QY 595 CATGTCATATGATCGGTGCGAGAAAGTCCGCATGATTAAACTGCACAGTGAAGAGT 654
DB 107 CATGCACACTGTCGCGGTGACGAGGTGATCTGTGAGAAATCCCGGAGCGGANGCT 48
QY 655 CTACTGGAGCAACGCCATAGCTTCCTATATACCCGCTATTC 699
DB 47 TGTTTTGCTCAACATGCTGGGCTCCCGCAAGCGCAATGGTG 4

Search completed: July 6, 2003, 06:40:28
Job time : 1978 secs

```

```

source
1..505
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC lPIG"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT      104 a 121 c 158 g 122 t
ORIGIN
Query Match      2.4%; Score 35; DB 12; Length 505;
Best Local Similarity 55.3%; Pred. No. 15;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 2 CCGGGAACTCCATGTGGCCGTGATCCTGGTCGAGCAGGATATIGCGATGATCCAGCGGC 61
DB 334 CCTCAATCTTCACCTGGTCTCAGTCTGTGCTCTGAAGGTGAATCGCGGTGCTTTTGGGAT 393
QY 62 CGCACAGCGCTGTGGGTATGATGAAGGCTGGTTGTAGAAACGCTGACCCCAACA 121
DB 394 TGAATACGTGGTGAGACCTGACCGTGGCCCTGCAGGCAGAGGAGCTGACCCCTCCAGCA 453
QY 122 GCT 124
DB 454 GCT 456

Search completed: July 6, 2003, 06:40:28
Job time : 1978 secs

```

## FEATURES

```

RESULT 15
BG834211
LOCUS      505 bp  mRNA  linear  EST 25-MAY-2001
DEFINITION 352345 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  BG834211
VERSION    BG834211.1  GI:14199231
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 505)
            Fabrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laedreid,W.W.
            and Keele,J.W.
            Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
            Unpublished (2000)
            Contact: Smith, TPI
            USDA, ARS, US Meat Animal Research Center:
            PO Box 156, Clay Center, NE 68933-0156, USA
            Tel: 402 762 4356
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 12
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTCACGTCACGAG
            Plate: 113  row: 1  column: 2
            Seq primer: ATTAGGTGACACTATAG.
            Location/Qualifiers

```

GenCore version: 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 15:36:56 : Search time 82 seconds  
(without alignments)  
533,002 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1745  
Sequence: 1 MWLFRESIMAKRGVAGKKP.....GMRGPKYIVGAMLNKLLV 324

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:
2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:
3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:
4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:
5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:
6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:
7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:
8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:
9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:
10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:
11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:
12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:
13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:
14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:
15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:
16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:
17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:
18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:
19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:
20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:
21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:
22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:
23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:

Pred. No. is the number of results predicted by change to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	99.0	328	19	AAW40263
2	355.5	20.3	3.5	21	AAW40263
3	233	13.3	296	23	ABW47924
4	208	11.9	432	23	ABW93276
5	198	11.3	452	21	AAW09207
6	198	11.3	452	21	AAW50180
7	178	10.2	404	21	AAW09208
8	178	10.2	404	21	AAW50181
9	177.5	10.2	575	19	AAW58856
10	174.5	10.0	400	21	AAW09209

11	174.5	10.0	400	21	AAW50182
12	148	8.5	33	19	AAW40275
13	130	7.4	307	23	ABW91277
14	117	6.7	21	19	AAW40272
15	107	6.1	238	23	ABW43158
16	105	6.0	842	23	ABW93909
17	101	5.8	525	22	AAW70539
18	101	5.8	525	22	AAW70540
19	101	5.8	525	22	AAW70541
20	101	5.8	526	22	AAW70542
21	101	5.8	629	23	AAE15854
22	101	5.8	853	23	AAE15853
23	101	5.8	993	22	AAW70537
24	101	5.8	994	22	AAW70538
25	101	5.8	994	22	AAW70543
26	100.5	5.7	261	22	AAW40273
27	100	5.7	23	19	AAW40274
28	99	5.7	20	19	AAW40264
29	98	5.6	18	13	AAW40270
30	94	5.4	19	19	AAW40271
31	93	5.3	682	21	AAW51947
32	91.5	5.2	194	22	AAW64716
33	91	5.2	1131	22	ABW24897
34	90	5.1	2748	22	ABW58843
35	89.5	5.1	635	22	AAU35532
36	88.5	5.1	428	23	ABW90560
37	88.5	5.1	443	20	AAW34467
38	88.5	5.1	451	20	AAW34343
39	88	5.0	773	23	ABW52004
40	87	5.0	850	22	AAW91859
41	86.5	4.9	428	20	AAW35396
42	86	4.9	388	22	ABW26871
43	86	4.9	1734	22	AAW20500
44	85.5	4.9	925	22	AAW92740
45	85	4.9	345	21	AAW47748

#### ALIGNMENTS

RESULTS  
AAW40263  
ID AAW40263 standard: Protein; 328 AA.

AC AAW40263  
ET 16-JUN-1998 (first entry)

DE K. oxytoca R-specific amidohydrolyase protein.

KW R-specific amidohydrolyase; hydrolysis; racemate; nitrogen source;  
KW 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.

OS Klebsiella oxytoca.

PK W09801568-A2.

XX 15-JAN-1996.

XX 10-JUL-1997; 97W0-EP03670.

XX 03-MAR-1997; 97CH-0000500.

XX 10-JUL-1996; 96CH-0001723.

XX (LANZ) LONZA AG.

XX Brieden W, Naughton A, Robins K, Shaw N, Tinschert A;

XX Zimmermann T;

XX WPI; 1998-101063/09.

XX N-PSDB; AAW10449.

XX Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation

PT - by stereoselective hydrolysis of corresponding racemic amide using  
 PT microorganism or derived enzyme, used as drug intermediate

PS Claim 5; Page 34-35; 58pp; German.

XX This sequence represents a R-specific amidohydrolase isolated from  
 CC Klebsiella oxytoca strain PRS1 which allows the microorganism to utilize  
 CC 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole nitrogen  
 CC source. This amidohydrolase is used in a process for preparing  
 CC (R)-isotoms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which  
 CC is cheaper than prior art optical resolution of the racemate using  
 CC dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.

XX Sequence 328 AA;

Query Match: 99.0%; Score 1731; DB 19; Length 328;

Best Local Similarity 99.4%; Pred. No. 7e-170; Indels 0; Gaps 0;  
 Matches 326; Conservative 0; Mismatches 2;

QY 1 MKWLEESIMAKRGVAGKRPVTHHTEEMOKERHYTIGPSTPVLTEPGDRIIVDTROA 60  
 DQ 1 MKWLEESIMAKRGVAGKRPVTHHTEEMOKERHYTIGPSTPVLTEPGDRIIVDTROA 60

QY 61 FEGAINSEQDIPSOQLKXKPELNPONGPIVNGAKGDVLAIVIESMLPRGVDPYG-CAMI 120  
 DQ 61 FEGAINSEQDIPSOQLKXKPELNPONGPIVNGAKGDVLAIVIESMLPRGVDPYG-CAMI 120

QY 121 PHFGGLTGDTLAMLNDPLPEKVRMIKLDSEKVKSKRHILPYKPHIGTILSVSPEDISIN 240  
 DQ 121 PHFGGLTGDTLAMLNDPLPEKVRMIKLDSEKVKSKRHILPYKPHIGTILSVSPEDISIN 240

QY 181 SLTPDNHGGKMDVPDIDGPGSITVLPVRAPGGRFLIGDAHACGGDGE-CGTAVFEFASITTI 240  
 DQ 181 SLTPDNHGGKMDVPDIDGPGSITVLPVRAPGGRFLIGDAHACGGDGE-CGTAVFEFASITTI 240

QY 241 KVDLIKXWOLSWPRMENAENIMSGSARPLEDATRIAYRDLIYWLVEDFGFEQWDAYNEL 300  
 DQ 241 KVDLIKXWOLSWPRMENAENIMSGSARPLEDATRIAYRDLIYWLVEDFGFEQWDAYNEL 300

QY 301 SQCGKVRGLGNKMDPKYTVGAMLNKNLLV 328  
 DQ 301 SQCGKVRGLGNKMDPKYTVGAMLNKNLLV 328

RESULTS 2

ID AAB29631 standard; protein; 315 AA.

AC AAB29631;

ET 23-FEB-2001 (first entry)

XX Enterobacter cloacae stereoselective amidase.

XX Amidase; stereoselective hydrolysis; alpha-amino acid amide;  
 KW alpha-hydroxy acid amide; optically active product;  
 KW recombinant production.

OS Enterobacter cloacae.

PN W0200053354-A1.

XX 26-OCT-2000.

XX 17-APR-2000; 2000WO-JP02492.

XX 16-APR-1999; 99JP-0109329.

XX (MITR) MITSUBISHI RAYON CO LTD.

XX Nakamura T, Yu F;

XX WPI; 2000-672731/65.

DR N-PSDB; AAC64148, AAC64149.

XX Novel amidase gene encoding protein; which stereoselectively hydrolyzes  
 PT alpha-amino-acid amides and alpha-hydroxy-acid amides, with enhanced  
 PT catalytic capability to yield optically active alpha-amino-acids and  
 PT alpha-hydroxy-acids

XX Claim 1; Page 20-22; 32pp; Japanese.

XX The invention relates to a novel amidase, and the gene encoding the  
 CC novel amidase, from Enterobacter cloacae. The novel amidase  
 CC stereoselectively hydrolyses alpha-amino acid amides and alpha-  
 CC hydroxy acid amides to give optically active alpha-amino acids and  
 CC alpha-hydroxy acids. The invention also relates to variants of the novel  
 CC amidase which retain activity, recombinant vectors and host cells  
 CC containing the novel amidase gene, and recombinant production of the  
 CC novel amidase. Genetically manipulated microorganisms comprising the  
 CC novel amidase gene have remarkably elevated catalytic capability as  
 CC compared to those produced by conventional methods. The present sequence  
 CC represents the novel Enterobacter cloacae amidase.

XX Sequence 315 AA;

Query Match 20.3%; Score 355.5; DB 21; Length 315;

Best Local Similarity 31.9%; Pred. No. 5.5e-28;

Matches 100; Conservative 42; Mismatches 136; Indels 35; Gaps 9;

QY 23 HHTEEMOKERHYTIGPSTPVLTEPGDRIIVDTRODAFEGAIN---SEQDIPSOQLKMP 79  
 DQ 23 HHTEEMOKERHYTIGPSTPVLTEPGDRIIVDTRODAFEGAIN---SEQDIPSOQLKMP 79

QY 80 F--LNPONGPIVNGAKGDVLAIVIESMLPRGVDPYGTCAMIPHFGGLTGDTLAMLND 137  
 DQ 80 F--LNPONGPIVNGAKGDVLAIVIESMLPRGVDPYGTCAMIPHFGGLTGDTLAMLND 137

QY 61 FDKLNPVSGPVVVEGAQPGDALKVTLSEFRPSG---FGWTANIPGFGLLAD-----QFSD 112  
 DQ 61 FDKLNPVSGPVVVEGAQPGDALKVTLSEFRPSG---FGWTANIPGFGLLAD-----QFSD 112

QY 138 PLPEKVRMIKLDSEKVKSKRHILPYKPHIGTILSVSPEDISINSLTPDNH-CGNMDVPDI 196  
 DQ 138 PLPEKVRMIKLDSEKVKSKRHILPYKPHIGTILSVSPEDISINSLTPDNH-CGNMDVPDI 196

QY 113 PALTUQYDROGLTPCAFGRYGRVPLKPFAGTLGVAPAAAGHHVYVPPRRVGGNLDIIDL 172  
 DQ 113 PALTUQYDROGLTPCAFGRYGRVPLKPFAGTLGVAPAAAGHHVYVPPRRVGGNLDIIDL 172

QY 197 GPGSITVPLVRAPGGRFLIGDAHACGGDGEICGTAVFEFASITTIKVDLIKXWOLSWPRME 256  
 DQ 197 GPGSITVPLVRAPGGRFLIGDAHACGGDGEICGTAVFEFASITTIKVDLIKXWOLSWPRME 256

QY 173 AAGCTLWLPVEVEGALFSGDTHAAGDGVHCGGTALIESANQVYVVKLVKDPKLTIPFA 232  
 DQ 173 AAGCTLWLPVEVEGALFSGDTHAAGDGVHCGGTALIESANQVYVVKLVKDPKLTIPFA 232

QY 257 NAENIMS-----IGSARPLEDATRIAYRDLIYWLVEDFGFEQWDAYNELLSQCGKV 306  
 DQ 257 NAENIMS-----IGSARPLEDATRIAYRDLIYWLVEDFGFEQWDAYNELLSQCGKV 306

QY 307 RLGNMVD-PKYYTV 318  
 DQ 307 RLGNMVD-PKYYTV 318

QY 293 RTSEIVDRPNVVV 305  
 DQ 293 RTSEIVDRPNVVV 305

RESULTS 3  
 ABA47924  
 ID ABA47924 standard; Protein; 296 AA.

XX ABA47924;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #625.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX W0200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

```

PR 11-APR-2000; 2000FR-0004623.
XX (INSP ) INST PASTER.
PA
XX
XX Buchrieser C, Franquel E, Courvo E, Rusnick C, Tsibi H, Dehoux P;
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Roland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Donnan E, Hain T, Berche P, Chastit A, Durand L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madenico E, De Pablos B, Wehland J, Kanst U, Estlin K, Hauf J;
PI Rose M, Voss H;
XX
XX WPI: 2002-G10914/01.
DR
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides
XX
XX Claim 6: SEQ ID NO 629; 132pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ARA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccines compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 296 AA;
SQ
Query Match 13.38; Score 233; DB 23; Length 296;
Best Local Similarity 25.08; Pred. No. 2.5e-5;
Matches 74; Conservative 62; Mismatches 116; Indels 24; Gaps 2;
QY 28 EMKKEPHYTIGPSTPVLTIEPGDRIIVDTAFEGAI--NSOODIPSOILKMPFLNPQ 87
DB 13 EMKKS-----TEPAIRVADGSSVVKIKRHNNGCHAKQLNHEJDMKQF-SPTTP 63
QY 38 IYVNGAKGRVLAVYIESMLPRGVDPYGTICAMIFHFGGLTGTSLTAMNDPFLPKVMTK 147
DB 64 IYIEARFGDLLATIEKIELLGTPEVFLNG--PNT-GITDMLTS-----NSTEQYK 155
QY 148 LDSKVVVSKRHTPYKPRHGTCTLSVSPEDSIN-SLTPNKHGNDVDPDIPGASITFLV 256
DB 114 VENQHIYSEDIPPIPKTKTCLL-----KTEELAFSKVPTKNGGLPSSKITREGATFDPY 169
QY 207 RAPGRRLFGDAHACGDGEIGTGAFFASITCKVPLIKNWOLSWHPRMENAKNINMSGS 266
DB 170 EKYGASLVGNVRATTFGKITATSAAPAEVTLRLQIAKNRATPTTIHYNLECLIAS 259
QY 267 ARPLEDATRIAYROLIYWLVEDFGEQWDAYMILASQGVKVRLONMVSPKYTVGAML 322
DB 230 DITTEKATOKTMHNNITLLIESDKWITFDALFLISLQADFOVCKLCKPKNITSIKL 285
RESULT 4
ID ABB93276
XX ABB93276 standard; Protein; 432 AA.
XX
XX AC ABB93276;
XX

```

---

```

PI 31-MAY-2002 (first entry)
XX
XX Herbicidally active polypeptide SEQ ID NO 2487.
XX
XX Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
XX W020210210-A2.
XX
XX 07-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-EP09892.
XX
XX 25-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidner M;
XX
XX WPI: 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5: SEQ ID NO 2487; 261pp + sequence listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 432 AA;
SQ
Query Match 11.98; Score 208; DB 23; Length 432;
Best Local Similarity 24.18; Pred. No. 1.8e-12;
Matches 92; Conservative 44; Mismatches 140; Indels 106; Gaps 16;
QY 28 EMKKEPHYTIGPSTPVLTIEPGDRIIVDTAFEGAI--NSOODIPSOILKMPFLNPQ 84
DB 20 QDQQLHNKHPPIPPVAEVKAGEFFRVEIDAMGGVINKDNDSDAIKN--LVLTTHHL 77
QY 85 NGPIMV-----NSAEKGVVLAVYIESMLPRGVDPYGTICAMIFHFGGLTGTSLTAMNDPL 139
DB 79 SGPIRVVDEEGVAAGKAGDLLAVEICNLGLPQDEMGFTGSPDRENG-----GGFLTDHF 131
QY 140 PEKVMKILDESKVYVSKHILPYK-----HIGTILSVSP----- 174
DB 132 P-----CATKAIWYFEGIYAVSPQIVGRFPGLTGHPGVIGTAPSNELLRIWNRERQ 183
QY 175 -EIDSTNSL-----PD-----NHGNDVDPDIPGASITFLVRA 210
DB 184 LEESGVESLTCEIEKGTPEWERANEARTINGRENGCNDIKW.SGSKYILPVPEV 243
QY 211 GRKFTGDHACGDGEI--CGTAVHFASITTKVPLIKNWOLSW-----PRMENAKNIMS 263
DB 244 ANLSTGDMHFSGDGEISFGCC-AIEMSGFLEKICBIIRNGHMOEVLTPMGPTPLHVNPIFE 302
QY 264 IGSARP-----LEDATRIAYROLIYWLVE---DFGEQWDAYMLL 300
DB 303 IGPVEPRFSEMLVFEGISVDESCKOYLDAT-VAYKRAVLNAIDYLFKEGYSKQOVILL 361
QY 301 SCCG-KVRLGNMVCPKYTVGAM 321
DB 362 SCCPCGRLSGIVDSPNAVTL 383

```



RESULT 5  
AAG09207  
ID AAG09207 standard; Protein: 452 AA.  
XX AC AAG09207;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 7049.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP:033405-A2.  
XX PD C6-SEP-2000.  
XX RF 25-FEB-2000; 2000EP-030-439.  
XX RF 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131149.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132488.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134222.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135622.  
PR 25-MAY-1999; 99US-0136022.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 05-JUL-1999; 99US-0142055.  
PR 05-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

```
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149925.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153738.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154773.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157763.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160985.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161353.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161952.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.3%; Score 198; DB 2; Length 452;
Best Local Similarity 24.9%; Pred. No. 2e+11;
Matches 100; Conservativity 46; Mismatches 131; Indels 124; Gaps 19;

QY 28 ENOKEFHTTIGYSTPVLITPEGDHIVDTRDAFCA-NSQPTPSQLIKKFFINPQ 84
DB 20 QQQQPLNRWRHPEIPPAEVKAGEFVEMIDAMGV-KDQSDASDKN-IVITTHHL 77
QY 85 NGPIWV-----NGAEKGDVLAVIESMLPRGVDPYGI-----CA---113
DB 78 SGPIRWDEEGVAARAGDOLJAVEICNLPLPGDEWGF-GSFDRENGGGHLDHFFPCATKA 137
QY 119 -----MIP--HFGGJT--GTDLTAMLNDPLPEKVRMIKLDSEKVIWSKRH---159
DB 138 IWYFESIVAYSQIPGVFRFPGLTHPGVIGTA---PSNELLRI-----WDRERQL 184
```

160 -----TLPYKPHIGTILSVSP-----IDSTNSLIPD-----NHGGM 191  
185 ERSVESLTCEVWHQRIACLPTRKGLIGNIEGTPEWERIANEAARTIPKRENGNC 244  
192 DVPDIGPGSTYPIVRAPGGRIFGDHACOGPEI--CGTAVEFASITIKVDLIKNNQ 249  
245 DINKLSRGSKIYLPVVEGANLSIGDHFSQDGEISFCG-AIEMSGFLEKKEIIRNGM 303  
250 LSW-----PRMENAENINSIGSARP-----LEDATRIAYRDLIYW 284  
304 QEVLTPMGPTPLHVNDFRIGVPVPRFSEWLVFEGTSVDSGKQHYLDAT-VAYKRAVLN 362  
285 LVE---GFGFEOWDAYMILSCCG-KVRLGNMWDPKYTVGAM 321  
363 AIDYLFKFGYSKEQVYLLLSCCPCFGRLSGIVDPSAVATL 403

RESULT 6  
AAG50180  
ID AAG50180 standard; Protein; 452 AA.  
XX  
AC AAG50180;  
XX  
XX 18-OCT-2000 (first entry)  
DX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63561.  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63561.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123380.  
PR 09-MAR-1999; 99US-0123546.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132437.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132643.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154775.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158024.  
PR 08-OCT-1999; 99US-0158212.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.

REFUGIT 7



XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123526.  
PR 23-MAR-1999; 99US-0125785.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130445.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134259.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134769.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 26-JUL-1999; 99US-0144894.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145036.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 25-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149366.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151303.  
PR 27-AUG-1999; 99US-0151304.  
PR 27-AUG-1999; 99US-0151305.  
PR 27-SEP-1999; 99US-0152363.  
PR 07-SEP-1999; 99US-0153070.  
PR 10-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155655.  
PR 28-SEP-1999; 99US-0156456.

PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159305.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161820.  
PR 28-OCT-1999; 99US-0161892.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 10.2% Score 178; DB 21; Length 404;  
Best Local Similarity 25.4%; Pred. No. 2e-09;  
Matches 94; Conservative 40; Mismatches 112; Indels 124; Gaps 19;  
QY 59 DAFEGAI---NSEODIPSGAKKPFPLNPONGPIWY-----NGAEKGDVLAAYIESMLPRG 110  
DB 3 DAMGGVTKDMSADIKN---LVLTTHLSPRYWDEEKVAKAGGGLJAVEICNIGLPF 60  
QY 111 VDPYGI-----CA-----MIP---RFGILT--GTHLTA 133  
DB 61 GDWGTGTSDFRNGSGCHLTHFFPCA-KA-WYFAGIVAYSQIFGVNFRPLTHFGVZGA 120  
QY 134 MLNDPLPEKVRMKDSEKIVYKSRH-----TLPYKPHIGTSLVSPE-----ID 177  
DB 121 -----PSNELLRI-----WNDERQLBESGVSLTCEVYVQRPLACPTTKGGLIG 167  
QY 178 SINSILPD-----NHGGMNVPTIGTGSILTYPVVRAPGGRLFGDARAO 222  
DB 168 NIECTFWESEIANEAARTIPRGNGCNDIKNLRSKILYLPVFEVGANLS-GMGFFSQ 227  
QY 223 GIGETI--CGTAVERASITTKVDLIKNNQLSW-----PRMENAEINISIGSARF----- 269  
DB 228 GDGEISFCG-AIEMSGFELEKCEIIRNGOEYLTTPKSTPLHVNPIFEIGPVEPRFSEML 286  
QY 270 -----LEDATSTAYRDLTYWIVE---DFGFEQWADYMLSCCG-KVRLSNW 311  
DB 287 VFEISVDESKOXYLDAT-VAYKRAVLNAIDYLFKFGYSKEOYVLLSCCPGEGRLSGT 345  
QY 312 VDPKTYTVGAM 321  
DB 346 VDSPNAVATL 355

RESULT 9  
AAW58856  
ID AAW58856 standard; Protein: 575 AA.  
XX

AAW58856;  
23-JUL-1998 (first entry)  
C. acidivorans gamma-lactamase protein.  
Gamma-lactamase; enantiomer; bicyclic lactam; temperature stable;  
2-azabicyclo(2.2.1)hept-5-en-3-one; therapeutic agent; antiviral;  
cardiac vasodilator.  
Comamonas acidivorans.  
WC9810075-A1.  
12-MAR-1998.  
01-SEP-1997; 97WO-GB02344.  
03-SEP-1996; 96GB-0018340.  
(CHIR-) CHIROSCIENCE LTD.  
Brown RC, Lee CS, Wisdom RA;  
WPI: 1998-193625/17.  
N-ESDR: AAV11459.  
Lactamase enzyme specifically hydrolyzing (+)-enantiomer of 2-azabicyclo(2.2.1)hept-5-en-3-one - isolated from Comamonas acidivorans or prepared by recombinant DNA technology. Used as, e.g. cardiac vasodilator  
Claim 5; Page 17-20; 28pp; English.  
This sequence represents a gamma-lactamase capable of hydrolyzing an enantiomer of the bicyclic lactam 2-azabicyclo(2.2.1)hept-5-en-3-one. This enzyme has at least 1 of the following features: greater than 85% retention of activity after incubating at 40 deg. C for 4 hours, or >30% activity after incubating at 60 deg. C for 4 hours, hydrolysis at an initial concentration of 100 g racemic lactam plus 300 ml buffer and proceeding to at least 30 (especially at least 98%) hydrolysis of the (+) lactam with <5% hydrolysis of the (-) lactam. The enzyme is much more temperature-stable than previously identified (-)-gamma-lactamases used for the same purpose. It also enables the bioresolution to be carried out at very high substrate/product concentrations. This enzyme has applications as a therapeutic agent e.g. as an antiviral agent or as a cardiac vasodilator.  
XX Sequence 575 AA;  
Query Match 10.2% Score 177.5; DB 19; Length 575;  
Best Local Similarity 24.0%; Pred. No. 3.8e-09;  
Matches 69; Conservative 33; Mismatches 85; Indels 101; Gaps 12;  
QY 39 PYSTP-----VLITPEGRRIIVDTDAFEGAI---NSEQDIPS-QLKMKPLN 82  
DB 44 PYDNPOVNRHNPIDIPMAVMEVPEGAFLIETDWTGSAIKNDGSAEDVCDMLSTVHFL- 102  
QY 83 PONGPIWNGAEKGDVLAAY-----YIESMLPRGV-----D 112  
DB 103 --SGPVGVGAOPGOLLVVDLLDLCARDOSLWGFNGFFSKONGGGLDEHFLAKSIND 160  
QY 113 PYGICAMIPH-----FGLTGTDLTAMLDPLPEKVRMKILDSEKIVYKSRHITLPYKPHI 167  
DB 161 PHGMFTKSRHIPGVNFAGLIHPGLIGCLPDP-----KML-----ASNRETI----- 202  
QY 168 GTLSVSPE-----IDINSILPDNHGGMNVDPDTPG 199  
DB 203 GLIATDPPRIPGLANPPNATTAHMQMOGEARDAKAAEGARTVPPRPHGCGNCDKLSRG 262  
QY 200 SITYPVLRAPGGRLFGIDAHACOGDGEIC-GTAVEFASITTKVDLIK 246  
DB 263 SRVFFPVYVDGAGLSVGLHFSQGDGEITFWNGPIEMPGVWHKVSLLK 310

RESULT 10  
AAG09209  
ID AAG09209 standard; Protein; 400 AA.  
XX  
AC AAG09209;  
XX  
D-  
XX  
DE 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7051.  
XX  
KW Protein identification; signal transduction; pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
DN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000: Z000EP-C301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123348.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130610.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134788.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139761.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 32-AUG-1999; 99US-0146386.  
PR 32-AUG-1999; 99US-0146388.  
PR 32-AUG-1999; 99US-0146389.  
PR 33-AUG-1999; 99US-0147038.  
PR 34-AUG-1999; 99US-0147204.  
PR 34-AUG-1999; 99US-0147302.  
PR 35-AUG-1999; 99US-0147192.  
PR 35-AUG-1999; 99US-0147260.  
PR 36-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 15-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.



```
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150888.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155149.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156438.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159273.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159431.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160983.
PR 22-OCT-1999; 99US-0160988.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161355.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161592.
PR 28-OCT-1999; 99US-0161893.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 10.0%; Score 174.5; DB 21; Length 400;
Matches 90; Conservative 41; Mismatches 113; Indels 121; Gaps 18;

QY 63 GAINSEQIPQL--LKKPFLNPQNGPIW-----NGAEKGVLAIVIESMLPFGVDPYG 115
DB 2 GGVINDSASDIKNLVLTTHHLSGPIRVYDFRGVAAKAGCLLAVETICNLGPPLGQFNG 61
QY 116 I-----CA-----MIP--HFGSLT--GTDLIAMINOF 138
DB 62 FTGSDRENGGFLTDHFCATKAIWFEIGIYVSPQPGVREFGIITHGVIGTA----P 117
QY 139 LPEKVM(KLDSKVKYWSKRH-----TLPKPHIGTLISVSPF-----TDSINSL 182
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150888.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155149.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156438.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159273.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159431.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160983.
PR 22-OCT-1999; 99US-0160988.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161355.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161592.
PR 28-OCT-1999; 99US-0161893.
PR 29-OCT-1999; 99US-0162142.

118 SNELLRI-----WDRERQLSEGVESLCLCEVWHORPLACLPTTKGCLLGNIEG 168
183 TPD-----NHCGNMVDPIGPGS*TYPLVRAPCGRLIFIGDAHACQGDGEI 227
169 TPWEIRIANEAARTIPGRENGGNCIDKNLSRGSKIVLPVFVECANLSTGDMHFSQGDGEI 228
228 --CGTAVEFASITTIKVDLINKNQLSW-----PKMENAENIMSIGSARP----- 269
229 SFCG-AIEMSGFELKEIIRNGMQHYLTPTMGPTPLHVNPIFEIGHVPRFSEWLVFEGI 287
270 -----LEDATRIAYRDIIVWLVE---DFGFEQWDAYMLLSQCG-KVRLGNMYDPKY 316
288 SVDESGKHVLDAT-VAYKRAVLNAIDYLFKFGYSKEOVYLLLSGCCPCGRLSGIVDSFS 346
317 TVGAM 321
347 AVATL 351

RESULT 1:
AAC50182 standard; Protein: 400 AA.
AAC50182;
18-OCT-2000 (first entry)
Arabidopsis thaliana protein fragment SEQ ID NO: 53563.
Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EF3333405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301435.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
23-APR-1999; 99US-0130891.
28-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134236.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
```





```

06      | | | | | | | | | | | | | | | |
       1 EFHYTIGPSTPVLTIEPDGR 21

RESULT 15
ABR49:58 standard; Protein; 238 AA.
XX AC AB549158;
XX ET 05-FEB-2002 (first entry)
XX DE Listeria monocytogenes protein #1862.
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX KW vitamin B12; bacterial infection; disease.
XX QS Listeria Monocytogenes.
XX PN WQ200177335-A2.
XX PD 18-OCT-2001.
XX PF 1J-APR-2001; 203LWO-PF01118. *
XX PP 1I-APR-2003; 2003PR-0064629.
XX PT (INSP ) INST PASTEUR.
XX PU Buchrieser C, Frangeul E, Couve E, Kushiok C, Fsihi H, Dehoux P,
PI Gussurgeat O, Chetouiati F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kseft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Carrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Doman E, Main I, Berche P, Charbit A, Durant L;
PI Perez-Stas J, Buquero F, Garcia del Portillo F, Gomez-Lopez N;
PI Macduenio E, DC Pablo B, Wehlard J, Kaerst U, Entian K, Hauf J;
XX PI Rose M, Voss H;
XX WP1: 2002-010914/01..

Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
and prevention of Listeria and related bacterial infections, and
related polypeptides -
XX CS C-arm 6; SEQ ID NO 1863; 192pp; French.
XX CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-c (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Seq-ence 238 AA;

Query Match 6.1%; Score 107; DB 23; Length 238;
Best Local Similarity 22.2%; Pred.No. 0.019;
Matches 55; Conservative 24; Missatches 75; Indels 94; Gaps 9;

07 34 HYTGTPVLTIEPDGRI--IVTDRTAGEAINSEQIPSOCLKKPFUNPONGPIMVN 91
| | | | | | | | | | | | | | | | | | | | | |

```

Db 51 HKLKSPKRLILTHAGDHIIGLIAVKAAPEAL- ----- VMIG 48  
QY 92 GAEKGDVLAVYIESMLPROVDPYGICAMIPHEGSGTDLZAMLXNDPLPEKVRMIKLPSE 15:  
Db 89 SREK-----LAVETKEIYAFEAQK-----LKGYSDRFLVRIQIT:KMGD 129  
QY 152 KYVWSKRHTLPYKPHITGLSVSPE:DSINSLTPDNHGGNNVDPIGPGSITYFLVRAFGG 21:  
Db 130 MV-----GSILI-----IDTPGHTPGSTGFFDEK--NG 155  
QY 212 RLFIGDAHACQGDGETCGTAVEFASITIKVDLIKHWOLSWPRMENAENIKSIGSARULE 271  
Db 156 HLFVGDLFOTRGGAICGE-----KRWLFFPFAMGSDWDLPTSLASAENLQ 200  
QY 272 --DARIA 277  
Db 201 LFDVTEIA 208

Search completed: July 3, 2003, 15:47:16  
Job time : 84 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw mode:

Run on: July 3, 2003, 15:45:52 : Search time 24 seconds  
(without alignments)  
37.181 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1748

Sequence: 1 MKWLPSTMAKXGVNAGKRP.....GNVDPKITYGAMNKNKLDV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 01

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177.5	10.2	575	US-08-922-865-2	Sequence 2, Appli
2	177.5	10.2	575	US-09-510-949-2	Sequence 2, Appli
3	91.5	5.2	1848	US-08-296-791-6	Sequence 6, Appli
4	91.5	5.2	1848	PCT-US95-10661A-6	Sequence 6, Appli
5	82.5	4.7	500	US-08-926-8429-64	Sequence 64, Appli
6	82.5	4.7	857	US-08-779-113-2	Sequence 2, Appli
7	82.5	4.7	958	US-08-583-5625-2	Sequence 2, Appli
8	82.5	4.7	1752	US-09-556-877-180	Sequence 180, App
9	82.5	4.7	1752	US-09-620-412C-180	Sequence 30, App
10	82	4.7	540	US-09-177-165A-30	Sequence 42, App
11	81.5	4.7	1170	US-08-476-062A-42	Sequence 42, App
12	81.5	4.7	1170	PCT-US96-0131A-42	Sequence 4, App
13	81	4.6	1342	US-07-976-895-4	Sequence 4, App
14	81	4.6	1342	US-08-473-119-4	Sequence 4, Appli
15	81	4.6	1342	US-08-475-352-4	Patent No. 513864
16	81	4.6	1343	US-08-475-352-4	Sequence 24, Appli
17	80	4.6	628	US-07-952-853-22	Sequence 22, Appli
18	80	4.6	528	US-08-914-848-22	Sequence 46, Appli
19	79.5	4.5	1544	US-09-413-814-45	Sequence 2, Appli
20	79	4.5	374	US-09-306-446C-2	Sequence 10, Appli
21	79	4.5	1342	US-08-484-438-9	Sequence 10, App
22	79	4.5	1711	US-08-368-822C-10	Sequence 7, Appli
23	79	4.5	1711	US-08-582-776C-10	Sequence 10, App
24	79	4.5	1711	US-08-434-831B-10	Sequence 7, Appli
25	78.5	4.5	375	US-09-171-337A-7	Sequence 4, Appli
26	78	4.5	383	US-08-391-916A-4	Patent No. 5258290
27	78	4.5	526	US-08-391-916A-4	

ALIGNMENTS

RESULT 1  
US-06-922-865-2  
: Sequence 2, Application US/08922865  
: Patent No. 6090616  
: GENERAL INFORMATION:  
: APPLICANT:  
: TITLE OF INVENTION: MICROBICAN™, LACTAMASE ENZYME OBTAINED  
: TITLE OF INVENTION: THEREFROM, AND THEIR USE  
: NUMBER OF SEQUENCES: 2  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Protein Release #1.0, Version #1.39 (EPO)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/922,865  
: FILING DATE:  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 575 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
US-08-922-865-2

Query Match 10.2% Score 177.5; DB 3; Length 575;  
Best Local Similarity 24.0%; Pred. No. 3e-11;  
Matches 99; Conservative 33; Mismatches 85; Indels 101; Gaps 12;  
QY 39 PYSTP-----VLTIEPGRIIVDTRDAFEGAI---NSEQDIPS-QLKMPFLN 82  
DB 44 PYDNPQVHNKWHDPIDMAVWVPEGAEFKLETVDWTGGAIKNDSDADRVLDSTVHFL- 102  
QY 83 PQNSPIWVNAEKSDYLA-----YIESMLPRGV-----D 112  
DB 103 --SGPVGVKGQSPDVLVLDLDIGARDSDINMGFNQFFSKQNGGGLDEHFFPLAKSIND 160  
QY 113 PYGICAMIPH-----EGG:GTGDTATLNDPIPEKVRMIKLDSEKYVWSKRITLPYKPHI 167  
DB 161 FHGMFTKSRHPIGVNFAGLIHFGLICLPDP-----KML-----ASNRET----- 202  
QY 168 GTLSVSPF-----DSINSLTPDNHGNMNDVPDIPGS 199  
DB 203 GLIATDPDRIPGLANPPNATIAHMGOMGEARDKAAAGARTVPPRHGNGCNDKIDLSRG 262  
QY 200 SITYPYLVARGKRLFTGDHACGGGEGIC-GTAVEFASITITIKVDLIK 246  
DB 263 SKVFFVYVVGAGISVCDLHFSQGDGEITFWGPIEMPGVGHMKVSLIK 310

26 77.5 4.4 1170 2 US-08-789-078-2 Sequence 2, Appli  
29 77.5 4.4 1170 2 US-08-752-633-2 Sequence 2, Appli  
30 77.5 4.4 1170 5 PCT-US95-04886-2 Sequence 2, Appli  
31 77 4.4 565 4 US-09-142-623-11 Sequence 11, Appli  
32 77 4.4 575 1 US-08-403-866-7 Sequence 7, Appli  
33 76.5 4.4 215 1 US-08-107-684B-10 Sequence 10, Appli  
34 76.5 4.4 215 1 US-08-107-684B-11 Sequence 11, Appli  
35 76 4.3 456 4 US-09-268-364-21 Sequence 21, Appli  
36 76 4.3 629 4 US-09-134-001C-4354 Sequence 4354, Ap  
37 76 4.3 1065 3 US-08-630-172-9 Sequence 9, Appli  
38 76 4.3 1065 4 US-09-375-419-9 Sequence 9, Appli  
39 76 4.3 1177 4 US-09-134-001C-5106 Sequence 5106, Ap  
40 76 4.3 1222 2 US-08-682-517-15 Sequence 15, Appli  
41 76 4.3 1252 2 US-08-682-517-9 Sequence 9, Appli  
42 76 4.3 1346 3 US-09-320-878-4 Sequence 4, Appli  
43 76 4.3 1346 4 US-09-105-537-37 Sequence 37, Appli  
44 76 4.3 1187 4 US-09-105-537-6 Sequence 6, Appli  
45 75.5 4.3 354 4 US-09-500-569-12 Sequence 12, Appli

## RESULT 2

US-09-510-949-2  
 : Sequence 2, Application US/09510949  
 : Patent No. 6423522  
 : GENERAL INFORMATION:  
 : APPLICANT:  
 : TITLE OF INVENTION: MICROORGANISM, LACTAMASE ENZYME OBTAINED  
 : FROM THE INVENTION: THEREFROM, AND THEIR USE  
 : NUMBER OF SEQUENCES: 2  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.3C (EPO)  
 : CURRENT APPLICATION NUMBER: US/09/510.949  
 : FILING DATE:  
 : PRIOR APPLICATION NUMBER: 08/922,865  
 : FILING DATE:  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 575 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-09-510-949-2

Query Match 10.2%; Score 177.5; DB 4; Length 575;  
 Best Local Similarity 24.0%; Pred. No. 3e-11;  
 Matches 69; Conservative 33; Mismatches 85; Indels 161; Gaps 12;  
 QY 39 PYSTP-----VLTIEGDRIVDTDAFEAI---NSEQIPSP-QLLKMFELN 82  
 DB 44 PYDNQVNRWHPDIPMAVVEPGAEFKLEYDTGTGAIKNDGSAEDVVDLSTVIFL- 202  
 QY 83 PONGIMVNGAEKGDVLA-----YTESKMPKSV-----D 1:2  
 DB 103 --SGPVGKQACQDPLLVYDLDIGARDSDLGWNGCFKSGONGGFIIDEHFLACKSMD 160  
 QY 113 PYGICAMIPH-----FGGLGTDLTAMLDPLPEKVRMILKDSKVVYVSKRELTPYKPH: 167  
 DB 161 FUGMTKSHHIVGVNFAGLIHPLGLGLDPP-----KVL-----ASNNREI----- 2:2  
 QY 168 GLTSLVSPK-----TCSNLTDPNHHGGMMDVPIELPG 199  
 DB 203 GLIATDPDRIPGIANPPNATTATMTCMOGEARQKAAEGARTVPPRGHGNCDTKLSRG 262  
 QY 206 SITYPLVRAPGGRFLFIGDAHACGGGEGIC-CTAVEFASITIKVGLAK 246  
 DB 263 SRVFPVYVVGAGLSVGLHPSGCGGLTFWGPTEMPGVVHKVSLIK 310

## RESULT 3

US-08-296-791-6  
 : Sequence 6, Application US/08296791  
 : Patent No. 6245337  
 : GENERAL INFORMATION:  
 : APPLICANT: St. Geme III, Joseph W.  
 : TITLE OF INVENTION: Haemophilus Adherence and Penetration  
 : NUMBER OF SEQUENCES: 9  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
 : STREET: 4 Embarcadero Center, Suite 3400  
 : CITY: San Francisco  
 : STATE: California  
 : COUNTRY: United States  
 : ZIP: 94111-4187  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/296.791  
 FILING DATE: 25-AUG-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Treccartin, Richard F.  
 REGISTRATION NUMBER: 31,801  
 REFERENCE/DOCKET NUMBER: A-59941/RF\*/RMS  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3249  
 TELEX: 910 277299  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1848 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 US-08-296-791-6

Query Match 5.2%; Score 91.5; DB 4; Length 1848;  
 Best Local Similarity 19.1%; Pred. No. 1.4;  
 Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 20;  
 QY 35 YTGIPYSTPVLIEPGD---RIIVDTDAFE-GAINSE-QDIPSOQLKMPFLNPONGPI 88  
 DB 17 YALPYTEALVRDQVQIFRDFAEKNGKFSVGATNVEVRKKNQSL-----GSA 67  
 QY 89 MVNGAEKGDVAVYIESMLPRGVDPY-----GICAMIPHFGLGTGDTLAMLNDP 138  
 DB 68 LPNGIPMIDFSDVDKRIATLVNPQYVGVGVKHSVNGVSEL--HFGNLNGN-----MNG 120  
 QY 139 LPEKVRMILKDSKVVYVSKRHTLP-----YKPHIGTSLVSPFID 177  
 DB 121 NAKSHROVSSSENYTYVEKNFFTENVTSTKEEQDAQRREDYTPRLDKFVT--EVA 178  
 QY 178 SINSITPDNHHGGMMDVPIGPGSIYTP-LVRAPGGRFLFIGDAHACGGGEGICGTAVEFAS 236  
 DB 179 PIEASTANNKGEYNSD-----KYPAFVRLSGTGFYKKGK-----KYQL 220  
 QY 217 ITTK---VDLKKWQLSWPME-----NAENIMSIGSNRPLE---DA 273  
 DB 221 ILTEKDKOGNLEARNMDVGGDNLDELGVNAYTYGTAGTPYKVNHNENGLIGFGNSKERHSDP 280  
 QY 274 TRIAYRD-----LIY-----WIV-----EDF-----GFQWDAYMLLSQ 302  
 DB 281 KGLISODPLINAVLGDGSGPLFVYDREKGNKWLFLGSDYDFWAGYNKKSQWENYI----- 335  
 QY 303 CGKVLGNWVDPKYTVGAMLNKN 325  
 DB 346 --KHEFAEKYQOYSAGSLHGSN 356

RESULT 4  
 US-08-296-791-6  
 : Sequence 6, Application US/08296791  
 : GENERAL INFORMATION:  
 : APPLICANT: Washington University, et al.  
 : TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
 : NUMBER OF SEQUENCES: 9  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
 : STREET: 4 Embarcadero Center, Suite 3400  
 : CITY: San Francisco  
 : STATE: California  
 : COUNTRY: United States  
 : ZIP: 94111-4187  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10661A
: FILING DATE: 16-AUG-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,791
: FILING DATE: 25-AUG-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Irecartin, Richard F.
: REGISTRATION NUMBER: 31,601
: REFERENCE/DOCKET NUMBER: PP-59941/RFT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEFAX: 910 277299
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1848 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: PCT-US95-10661A-6

Query Match 5.2%; Score 91.5; DB 5; Length 1848:
Best Local Similarity 29.1%; Pred. No. 1.4;
Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 27;

QY 35 YTIQSTPVLTLTEPGD-----KIIVDTRDAFE-GAINSE-QYIPSQLKXPLFNTQNEP 88
DB 17 YALTPYTEAALVRDDVDYQ:FRDFAENKGFSGVATNVEVRKKNQS:-----GSA 67
QY 89 MYNGAEKGDVLAIVIESMLPRGVDPY-----GICAK:PIFGGLTGDLTAMLNDP 138
DB 68 LPLNGIPMDIFSVDVDEKRIATILNPQYVGVKHSNGVSEL--HFGNLGN-----MNG 120
QY 139 LPEKVRMIKIDSEKVMKRRHTLP-----YKPHIGTSLSVSPRID 177
DB 121 NAKSHRDVSSSENNRYTYVEKNFPTENTSTFKEDQAKRREDYMPRLXEVTV--EVA 178
QY 178 STNSLTPDNHGNMVDPIGPGSIYTP-LVRAPGRLFLIGAHACGGGECGTAVEFAS 236
DB 179 PIEASTANNKGEYNNSD-----KYPAFVRLGSGTQFYKKGS-----RYQL 220
QY 237 ITTIK---VDLIKWLSPRME-----NAENIMTSGSARPLE---DA 273
DB 221 ILTEKDKGNLLRNWDVGDGDLNLYGNAYTYGIAGTPYKVRHNNGLIGFNSKSEISDP 280
QY 274 TRIAYRD-----LIY-----WCV---EDF-----GPFQWDAYMLLSQ 302
DB 281 KGIISQDPLTNYAVIGDSGSPLEFYVDREKGGKFLGSDYDFWAGYNKKSQEWNIY---- 335
QY 303 CGKVRLGNNMVDPKYTVGAMLNKN 325
DB 336 --KHEFAEKIYQOYSAGSIIIGSN 356

RESULT 5
US-08-926-842B-64
: Sequence 64, Application US/08926842B
: Patent No. 6030807
: GENERAL INFORMATION:
: APPLICANT: Sa-No. 6030807uira, Isabel
: APPLICANT: de Lencastre, Herminia
: TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
: TITLE OF INVENTION: EXPRESSION
: NUMBER OF SEQUENCES: 64
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601

```

```

:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/926,842B
: FILING DATE: 10-SEP-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 603-1-089 N
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 64:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 500 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Bacillus subtilis
: FEATURE:
: OTHER INFORMATION: /product= "abfa"
: US-08-926-842B-64

Query Match 4.7%; Score 82.5; DB 3; Length 500:
Best Local Similarity 25.4%; Pred. No. 1.7;
Matches 5; Conservative 32; Mismatches 75; Indels 43; Gaps 13;

QY 142 EKVMTIKLUSE-----KVYMSKRHTLPYKPHGTLSVSPEDISINSIPDNHGNMD 192
DB 2 KRAMI-VDKEYKIGVDKRIYGS-----FIEHMR-AVYEGIVEPORPEADEGFRKD 53
QY 193 VPDGPGSTYDLVRAPGRLFLIGAHACGGGECG--TAVEFASITIKVOLKNNQL 250
DB 54 VQSL-IKELOVPIIRYPGNG-FLSGYNWEDGVGVENRPRRLDLAWOTETNEVGTNEFL 111
QY 251 SWPRMENAENINSIG-SARPLEDATRIAYRDLI-----YW--LVEDFGFEQWDAYM 298
DB 112 SWAKKVNTENVNNAVNIGIRGID-----AARNLVEYCNHPKGSYWSDLRRSHGYEQ--PYG 164
QY 299 LLSQCKVRLGNMVDPKYTVG 319
DB 165 IRTWC-----IGNEMOGPMQIG 181

RESULT 6
US-08-779-113-2
: Sequence 2, Application US/08779113
: Patent No. 5948891
: GENERAL INFORMATION:
: APPLICANT: Staunton, Donald E.
: APPLICANT: Harris, Edith S.
: TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
: TITLE OF INVENTION: Binding
: NUMBER OF SEQUENCES: 65
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gorstein, Murray & Borun
: STREET: 233 South Wacker Drive, 6300 Sears Tower
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:

```



APPLICATION NUMBER: US/08/779,113  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Greta E. No. 59489and  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/33773  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0446  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 857 amino acids  
TYPE: amino acid  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-779-113-2

Query Match 4.7%; Score 82.5; DB 2; Length 857;  
Best Local Similarity 20.3%; Pred. No. 4.3;  
Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 EESIMAKR-----GVGAGRKPVTHTLTEMOKEFHYTIGPYSTPVLTIPEGDR----- 52  
DB 204 DEPCLLKRLPNHHIGISFIPREVGELHVSIKKNGNHVANSVSYVQSEIGDARRAKVY 263  
QY 53 -----IIVDTRDAFEGAINSEODIPSQLKMPFLNPONGPIMVNGAEKGDV 98  
DB 264 GRGLSEGRTEFMSDFIVDTRDAGYGGIS-----LAVEGSPSKVDI 302  
QY 99 LAVVIESMLPRGVDPYIGICAMIPHFGLGTGDL--TAMLNDPLPEKVRMIKLDSEKVVYS 156  
DB 303 QTEPLED-----GTC-KVSYFTVPGVYVSTKFADEHVPGPSFTVVKISSEG--- 348  
QY 157 KRHTLPYKPHIGTLSVSPEDSINSILTPDNHGGNMDVPGISGITYPLVRAPGGRL 213  
DB 349 -----RVKESIRTSRAPSVATVGSIC----DLNLKIPETNSSDMS-AHVTSPSGRV 395

RESULT 7  
US-08-583-562B-2  
Sequence 2, Application US/08583562B  
Patent No. 5922570  
GENERAL INFORMATION:  
APPLICANT: Staunton, Dona'd  
APPLICANT: Harris, Edith  
TITLE OF INVENTION: Cytoplasmic Modulators of Integrin  
TITLE OF INVENTION: Binding  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 5300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/583,562B  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/33033  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0446  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 858 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-583-562B-2

Query Match 4.7%; Score 82.5; DB 2; Length 858;  
Best Local Similarity 20.3%; Pred. No. 4.3;  
Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 EESIMAKR-----GVGAGRKPVTHTLTEMOKEFHYTIGPYSTPVLTIPEGDR----- 52  
DB 204 DEPCLLKRLPNHHIGISFIPREVGELHVSIKKNGNHVANSVSYVQSEIGDARRAKVY 263  
QY 53 -----IIVDTRDAFEGAINSEODIPSQLKMPFLNPONGPIMVNGAEKGDV 98  
DB 264 GRGLSEGRTEFMSDFIVDTRDAGYGGIS-----LAVEGSPSKVDI 302  
QY 99 LAVVIESMLPRGVDPYIGICAMIPHFGLGTGDL--TAMLNDPLPEKVRMIKLDSEKVVYS 156  
DB 303 QTEPLED-----GTC-KVSYFTVPGVYVSTKFADEHVPGPSFTVVKISSEG--- 348  
QY 157 KRHTLPYKPHIGTLSVSPEDSINSILTPDNHGGNMDVPGISGITYPLVRAPGGRL 213  
DB 349 -----RVKESIRTSRAPSVATVGSIC----DLNLKIPETNSSDMS-AHVTSPSGRV 395

RESULT 8  
US-09-556-877-180  
Sequence 180, Application US/09556877  
Patent No. 6432916  
GENERAL INFORMATION:  
APPLICANT: Probst, Peter  
APPLICANT: Bhatia, Ajay  
APPLICANT: Skeiky, Yasar  
APPLICANT: Fling, Steve  
APPLICANT: Maisonneuve, Jeff  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C5  
CURRENT APPLICATION NUMBER: US/09/556,877  
CURRENT FILING DATE: 2000-04-19  
NUMBER OF SEQ ID NOS: 305  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 180  
LENGTH: 1752  
TYPE: PRI  
ORGANISM: Chlamydia  
US-09-556-877-180

Query Match 4.7%; Score 82.5; DB 4; Length 1752;  
Best Local Similarity 23.1%; Pred. No. 14;  
Matches 49; Conservative 26; Mismatches 60; Indels 77; Gaps 10;

QY 40 YSTPVLTIPEGDRRIIVDTRDAFE--GAINSEODIPSQLKMPFLNPONGPIMVNGAE--- 94  
DB 271 YAKGDLTIADSORVLV-SINKATKDGGAIPAERDV-----SFENITSLKVOINGAEK 323  
QY 95 -----KGDVL-----AVIESMLPRGVDPYIGICAMIPHFGLGTGD 130  
DB 324 GAIYAKGDLISIOSSKUSFNSNYSKGGGALYVE-----GGINFQD 364  
QY 131 LTAMLNDPLPEKVRMIKLDSEKVRHTLPKPHIGTISVSPEDSINSILTPDNHGGN 190  
DB 365 L-----EELR-IKYNKAGTTEIKTIL---PSLKAQASAGNADAWASSPQSGSA 411  
QY 191 MDVPDIPGPGSI-----TYPLVRAPGGRL 214  
DB 422 TTVSDSGDSSGSDSDTSETVP-VTAKGGGLY 442

RESULT 9

```
US-09-620-412C-180
: Sequence 180, Application US/09620412C
: Patent No. 6448234
: GENERAL INFORMATION:
: APPLICANT: Steven P. Fling
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: FILE REFERENCE: 210121.469C7
: CURRENT APPLICATION NUMBER: US/09/620.412C
: CURRENT FILING DATE: 2000-07-20
: NUMBER OF SEQ ID NOS: 363
: SOFTWARE: FASTSEQ for Windows version: 3.0/4.0
: SEQ ID NO 180
: LENGTH: 1752
: TYPE: PRT
: ORGANISM: Chlamydia
US-09-620-412C-180

Query Match          4.7%  Score 82.5; DB 4: Length 1752;
Best Local Similarity 23.1%; Pred. No. 14;
Matches 49; Conservative 26; Mismatches 60; Indels 77; Gaps 10;

QY 40 YSTPLVTIEPDRIVOTRDAFE--GAINSEODIPSSLLKMPFLNPONGPIVNGAF--- 94
DB 271 YAKGLTIADCEVLFSSINKAKKGGALFAEKW-----SFNTISLAVQNGASEKG 123
QY 95 -----KGNVL-----AVYIESMLPRGVWVPG:CAMIPIHFGGLGTG 130
DB 324 GATYAKGRLSOSSKQSLFNSYKQGGALYF-----GQINEQ 364
QY 131 LTAMLNPLPKVAMKLDSEKVVSKRHLLPYKPHIGTILSVAFEDSINSLTFDHNGGN 140
DB 365 L-----SEIR-IKYNKAGTETTKITL---PS:KAAGAGNADAKWASSSTFGSSGA 411
QY 191 MDVPDIPGSSI-----TYPELVKAPGGR:F 214
DB 412 TVSDSGSDSSGSDSDTSETVP-VTAKGGSIY 442

RESULT 10
US-09-177-165A-30
: Sequence 30, Application US/09177165A
: Patent No. 6426205
: GENERAL INFORMATION:
: APPLICANT: Tyets, Mike
: APPLICANT: Willemis, Andrew
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
: FILE REFERENCE: 11757.103US1
: CURRENT APPLICATION NUMBER: US/09/177.165A
: CURRENT FILING DATE: 1998-10-22
: PRIOR FILING DATE: 1998-07-10
: PRIOR APPLICATION NUMBER: 60/092,443
: PRIOR FILING DATE: 1998-07-10
: PRIOR APPLICATION NUMBER: 60/063,254
: PRIOR FILING DATE: 1997-10-24
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 30
: LENGTH: 640
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-30

Query Match          4.7%  Score 82; DB 4: Length 640;
Best Local Similarity 21.1%; Pred. No. 3;
Matches 39; Conservative 23; Mismatches 65; Indels 56; Gaps 7;

QY 62 NPONGPIVNGAEKGDVLAVYIESMLPRGVD:YGICAM:PIFGSLTGTGDTAM:KDLPL 141
DB 490 SPQDPTMTDGADESDTSPNEQTVLDR---NLPVPT 523
QY 142 KYRMKILJSEKVVYKSRHTLIPYKPHIGTILSVSPEDSINSLTFPN-----HCGNMDVP 194
```

```
DB 524 HLLSGCLDNTIKJWVKTGKCIQTQF-----HVEGVNDIAADNFRIISGSHDGIKW 577
QY 195 DIGPGSITVPLVRAPGGRL-----FIGD-----AHACQGDGEICGTAVEFASIT 239
DB 578 DLQSGKCHIF---NGRRLQRETOHTQSLGDKVAPIACVCIGDSE-CFSGDEFCVKM 633
QY 240 IKVDL 244
DB 634 YKFOL 638

RESULT 12
US-08-476-062A-42
: Sequence 42, Application US/08476062A
: Patent No. 5877275
: GENERAL INFORMATION:
: APPLICANT: Arnaud, M. Amlin
: TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
: TITLE OF INVENTION: RESPONSES WITH BETAZ INTEGRINS
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476.062A
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/216.081
: FILING DATE: 21-MAR-1994
: APPLICATION NUMBER: 07/637.830
: FILING DATE: 04-JAN-1991
: APPLICATION NUMBER: 07/539.842
: FILING DATE: 18-JUN-1990
: APPLICATION NUMBER: 07/212.573
: FILING DATE: 28-JUN-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Freeman, John W.
: REGISTRATION NUMBER: 29,066
: REFERENCE/DOCKET NUMBER: 00786/068003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 42:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1170 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: US-08-476-062A-42

Query Match          4.7%  Score 81.5; DB 2: Length 1170;
Best Local Similarity 22.7%; Pred. No. 9.4;
Matches 70; Conservative 35; Mismatches 109; Indels 95; Gaps 18;

QY 14 VAGAKKPYTHHLEKQKEFH-YIIGPYSTVPLTIEPCDRIIVTROAFEGAINSEQDIP 72
DB 284 IGIGK-----HFKTESQETLHKFKASKPASE-----FVKILOIFE----- 318
QY 73 SOLLKMPFLNPONGPIVNGAEKGDVLAVY:E-----SMLPRGVDPYIGICAM:PIHFG 125
DB 319 --KLDFLPIEROKKIYVIEGTSSKQDLTSFNMELSSSGISADLSRGHAVVGAVGAKWAGG 376
```



Db 1087 MPRGCLASSESGHVTGSEA-----ELQKVMCMCRSRSRSPRGRGDSAYHSORHSL 1140  
QY 162 -PYKPHICTLSVSP---EIDSINS-LTPDNH 187  
Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

RESULT 14  
US-08-473-119-4  
: Sequence 4, Application US/08473119  
: Patent No. 5820859  
: GENERAL INFORMATION:  
: APPLICANT: Kraus, Matthias H.  
: APPLICANT: Aaronson, Stuart A.  
: TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE  
: TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND  
: TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Suite 400  
: STREET: 133 Carnegie Way, N.W.  
: CITY: Atlanta  
: STATE: Georgia  
: COUNTRY: U.S.A.  
: ZIP: 30303  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/473,119  
: FILING DATE: 07-JUN-1995  
: CLASSIFICATION: 424  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/978,895  
: FILING DATE: 10-NOV-1992  
: APPLICATION NUMBER: US 07/444,405  
: FILING DATE: 01-DEC-1989  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Perryman, David G.  
: REGISTRATION NUMBER: 33,438  
: REFERENCE/DOCKET NUMBER: 1414-028  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (404) 688-9870  
: TELEFAX: (404) 688-9880  
: INFORMATION FOR SEQ ID NO: 4:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1342 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-473-119-4

Query Match 4.6%; Score 81; DB 2; Length 1342;  
Best Local Similarity 24.6%; Pred. No. 13;  
Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY 8 IMAKRGVAGRPVTHLTEMOKEFHYTIGPYSTPV-TIEPGDRIIVDTKAEFGAINS 67  
Db 978 IKRESGPGIAPGPEPHGLTNKKLEEVE--LEPELDLDLDEAEDNLATT-----TLGS 1029  
QY 68 EQDIPSOLLKMP-----FNPONGPIMVNGAEKGVAVYIESML-----PRGVD--- 112  
Db 1030 ALSLPVGTNNRPGSSLLSPSGYMPMNOQNLGSCQ---ESAVSGSSERCPRPVSLHP 1086  
QY 113 -PYGICAMIPHFGLTGTLATLNDPLPEKVRMKLDS-----EKVYWSKRHTL- 161  
Db 1087 MPRGCLASSESGHVTGSEA-----ELQKVMCMCRSRSRSPRGRGDSAYHSORHSL 1140  
QY 162 -PYKPHICTLSVSP---EIDSINS-LTPDNH 187  
Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

Search completed: July 3, 2003, 15:50:45  
Job time: 38 secs

Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

RESULT 15  
US-08-475-352-4  
: Sequence 4, Application US/08475352  
: Patent No. 5916755  
: GENERAL INFORMATION:  
: APPLICANT: Kraus, Matthias H.  
: APPLICANT: Aaronson, Stuart A.  
: TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE  
: TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND  
: TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Suite 400  
: STREET: 133 Carnegie Way, N.W.  
: CITY: Atlanta  
: STATE: Georgia  
: COUNTRY: U.S.A.  
: ZIP: 30303  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/475,352  
: FILING DATE:  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/978,895  
: FILING DATE:  
: APPLICATION NUMBER: US 07/444,406  
: FILING DATE: 01-DEC-1989  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Perryman, David G.  
: REGISTRATION NUMBER: 33,438  
: REFERENCE/DOCKET NUMBER: 1414-028  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (404) 688-0770  
: TELEFAX: (404) 688-9880  
: INFORMATION FOR SEQ ID NO: 4:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1342 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-475-352-4

Query Match 4.6%; Score 81; DB 2; Length 1342;  
Best Local Similarity 24.6%; Pred. No. 13;  
Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY 8 IMAKRGVAGRPVTHLTEMOKEFHYTIGPYSTPV-TIEPGDRIIVDTKAEFGAINS 67  
Db 978 IKRESGPGIAPGPEPHGLTNKKLEEVE--LEPELDLDLDEAEDNLATT-----TLGS 1029  
QY 68 EQDIPSOLLKMP-----FNPONGPIMVNGAEKGVAVYIESML-----PRGVD--- 112  
Db 1030 ALSLPVGTNNRPGSSLLSPSGYMPMNOQNLGSCQ---ESAVSGSSERCPRPVSLHP 1086  
QY 113 -PYGICAMIPHFGLTGTLATLNDPLPEKVRMKLDS-----EKVYWSKRHTL- 161  
Db 1087 MPRGCLASSESGHVTGSEA-----ELQKVMCMCRSRSRSPRGRGDSAYHSORHSL 1140  
QY 162 -PYKPHICTLSVSP---EIDSINS-LTPDNH 187  
Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

Search completed: July 3, 2003, 15:50:45  
Job time: 38 secs

-----

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Ruf on: July 3, 2003, 15:48:47 : Search time 53 seconds  
(without alignments)  
711.654 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1748  
Sequence: 1 MKWLESIMAKRGVAGRRK.....GMVLPKTVGAMLNKKLLV 329

Scoring table: ALOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 3  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US04\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/RCR\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCUS\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US63\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US63\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1748	100.0	328	9	US-10-086-082-2
2	151	9.2	416	9	US-10-156-761-9482
3	148	8.5	33	9	US-10-086-082-14
4	117	6.7	21	9	US-10-086-082-11
5	101	5.8	525	9	US-10-004-551-18
6	101	5.8	525	9	US-10-004-551-20
7	101	5.8	525	9	US-10-004-551-22
8	101	5.8	526	9	US-10-004-551-24
9	101	5.8	993	9	US-10-004-551-14
10	101	5.8	994	9	US-10-004-551-16
11	100.5	5.7	261	9	US-10-004-551-26
12	100	5.7	23	9	US-10-086-082-12
13	99	5.7	20	9	US-10-086-082-3
14	98	5.6	18	9	US-10-086-082-9
15	96	5.5	1094	9	US-10-086-082-11435
16	94	5.4	19	9	US-10-086-082-10
17	93	5.3	682	10	US-09-798-751-2
18	92.5	5.3	381	9	US-10-128-714-8002
19	91.5	5.2	1846	9	US-09-839-956-6

20	91.5	5.2	1848	9	US-10-080-505-6	Sequence 6, Appli
21	89.5	5.1	635	10	US-09-815-242-11225	Sequence 11225, A
22	87	5.0	850	9	US-09-738-626-5613	Sequence 5613, Ap
23	87	5.0	1907	9	US-09-832-292-39	Sequence 39, Appl
24	86.5	4.9	3338	9	US-10-156-761-8464	Sequence 8464, Ap
25	85.5	4.9	925	9	US-09-738-626-6494	Sequence 6494, Ap
26	84.5	4.8	747	10	US-09-815-242-5077	Sequence 5077, Ap
27	84	4.8	355	9	US-10-128-714-3002	Sequence 3002, Ap
28	83.5	4.8	1083	9	US-09-738-626-6914	Sequence 6914, Ap
29	83	4.7	513	10	US-09-833-745-58	Sequence 58, Appl
30	82.5	4.7	755	9	US-09-738-626-3810	Sequence 3810, Ap
31	82.5	4.7	826	9	US-09-746-660A-92	Sequence 92, Appl
32	82.5	4.7	833	9	US-09-746-660A-90	Sequence 90, Appl
33	82.5	4.7	1221	10	US-09-738-626-5163	Sequence 5163, Ap
34	82.5	4.7	1221	10	US-09-910-891-2	Sequence 2, Appli
35	82.5	4.7	1752	10	US-09-841-132-180	Sequence 180, App
36	82	4.7	640	9	US-10-060-019-30	Sequence 30, Appl
37	81.5	4.7	580	9	US-10-156-761-8637	Sequence 8537, Ap
38	81	4.6	652	10	US-09-815-242-10449	Sequence 10449, A
39	81	4.6	677	10	US-09-815-242-12100	Sequence 12100, A
40	81	4.6	1342	9	US-10-172-620-16	Sequence 16, Appl
41	80.5	4.6	395	9	US-09-738-626-6644	Sequence 6644, Ap
42	80.5	4.6	478	10	US-09-841-132-389	Sequence 389, App
43	80.5	4.6	520	9	US-10-156-761-11162	Sequence 11162, A
44	80	4.5	2503	9	US-10-007-706-1	Sequence 1, Appli
45	79.5	4.5	583	9	US-10-156-761-11791	Sequence 11791, A

## ALIGNMENTS

RESULT 1	
US-10-086-082-2	
Sequence 2, Application US/10086082	
Publication No. US20030687402A1	
GENERAL INFORMATION:	
APPLICANT: Riodes, Walter	
APPLICANT: Naughton, Andrew	
APPLICANT: Robins, Karen	
APPLICANT: Shaw, Nicholas	
APPLICANT: Tinschert, Andreas	
APPLICANT: Zimmermann, Thomas	
TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)	
TITLE OF INVENTION: -3,3,3-TRIFLUORO-2-HYDROXY-2-METHYLPROPIONIC ACID	
FILE REFERENCE: 32213	
CURRENT APPLICATION NUMBER: US/10/086,082	
CURRENT FILING DATE: 2002-02-28	
PRIOR APPLICATION NUMBER: US/09/214,679	
PRIOR FILING DATE: 1999-12-30	
NUMBER OF SEQ ID NOS: 14	
SOFTWARE: FastSeq for Windows Version 3.0	
SEQ ID NO 2	
LENGTH: 328	
TYPE: PPT	
ORGANISM: Klebsiella oxytoca	
US-10-086-082-2	
Query Match	100.0%; Score 1748; DB 9; Length 328;
Best Local Similarity	100.0%; Pred. No. 5e-163;
Matches 328; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MKWLESIMAKRGVAGRRKVTHTLTEMQKEFHYYTIGPSTVLTIFPDRIIVDTDA 60
Dh	1 MKWLESIMAKRGVAGRRKVTHTLTEMQKEFHYYTIGPSTVLTIFPDRIIVDTDA 60
QY	61 PEGAINSEQDIPSOQLKMPFLNPNQNGPIMVNGAKGDLVAVYTESM:PGVDVPGICAMT 120
25	61 PEGAINSEQDIPSOQLKMPFLNPNQNGPIMVNGAKGDLVAVYTESM:PGVDVPGICAMT 120
QY	121 PFGGLTCTDITAMLNQPLPEKVRMIKIDSEKVIWSKRHITLPYKPHIGTILSVSPEDSIN 180
145	121 PFGGLTCTDITAMLNQPLPEKVRMIKIDSEKVIWSKRHITLPYKPHIGTILSVSPEDSIN 180

```
QY 181 SLTPDNHGGNDVDPDTPGSSITYPLVRAPGRLPFGDAHACGGRGECGTAFAFASITL 240
DB 181 SLTPDNHGGNDVDPDTPGSSITYPLVRAPGRLPFGDAHACGGRGECGTAFAFASITL 240
QY 241 KVDLIKNNQLSWPRMNAENIMSGSARPLFDATEFAYRCLLYWLVECHGEQWDAFMLE 400
DB 241 KVDLIKNNQLSWPRMNAENIMSGSARPLFDATEFAYRCLLYWLVECHGEQWDAFMLE 400
QY 301 SOCGKVRLLGNVDPKPYTGAMINKNLLV 328
DB 301 SOCGKVRLLGNVDPKPYTGAMINKNLLV 328
RESULT 2
US-10-156-761-9482
; Sequence 9482, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/20156761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272597
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9482
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9482
Query Match
Best Local Similarity 25.2%; Score 161; DB 9; Length 416;
Matches 91; Conservative 42; Mismatches 122; Indels 106; Gaps 18;
QY 44 VLIIFGDRITVDTRDAFF---GAINSEQDTPSGLKXPFINTGNGP-KVNGAKKQDVLA 160
DB 32 VAMVKPGAEPRECRDNTDCQVGNKDTANDVRDIDELIPHM--LSGSGVGEAGPGLLV 99
QY 101 VVIESMLP-----RGV-----DPYGICAMIPHF 123
DB 90 VDIIDLGVPGQTDGAAGQGWYTGIFAKANGSGFLTDYFFDAYKAYWDER-IGQAVSRHL 249
QY 124 GGLTGIDLT-AHLNDPLPEKVRMKLDSEKVVYSKRH-----THPYKPHGISTLSVSPED 177
DB 150 PGIRFGITGTHPLGTAPSAEMLAR-----WTKRQALIDIDUFNR--VFPLAVTP--D 159
QY 178 SINSL-----TPD-----NHGNGVVEFDIGRSTITYPLVRAPGRLPFG 217
DB 199 ATNALACTATGDLARRTGEAGTVPARENGGNEIDKNTGSAFVYFVHWQAKLGGD 258
QY 218 ARACGGTGEI--CCTAVEFAFASITIKVDIKNNQLSW-----PRMNAENI 261
DB 259 LHFSGDGEITFCG-AIEMKGFIDFVDLIKGGMETYGISTNPVF-IPGNVEHYTETLIF 317
QY 262 MSIGSARPLE-----DATFAYRCLLYWLVE--DFGFEQWCAFMLE--SCGKVRKGNMV 312
DB 318 LGISVDHDTDINYLDAT-LAYRACLNAVYEFKKFGYSGEQAYLLGSSPLERISGIV 376
QY 313 C 313
DB 377 D 377
```

```
RESULT 3
US-10-086-082-14
; Sequence 14, Application US/10086082
; Publication No. US20030087402A1
; GENERAL INFORMATION:
; APPLICANT: Brieder, Walter
; APPLICANT: Naughton, Andrew
; APPLICANT: Robins, Karen
; APPLICANT: Shaw, Nicholas
; APPLICANT: Zimmermann, Thomas
; TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)
; FILE REFERENCE: 3,3,3-TRIFLUORO-2-HYDROXY-2-METHYLPROPIONIC ACID
; CURRENT APPLICATION NUMBER: US/10/086,082
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US/09/214,679
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Klebsiella oxytoca
US-10-086-082-14
Query Match
Best Local Similarity 8.5%; Score 148; DB 9; Length 33;
Matches 29; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 110 GVDPYGICAMIPHFGLTGTLTAMINDPLPEK 142
DB 1 GVDPYGIEAMIPHFGLTGTLTAMINDQLQPK 33
RESULT 4
US-10-086-082-11
; Sequence 11, Application US/10086082
; Publication No. US20030087402A1
; GENERAL INFORMATION:
; APPLICANT: Brieder, Walter
; APPLICANT: Naughton, Andrew
; APPLICANT: Robins, Karen
; APPLICANT: Shaw, Nicholas
; APPLICANT: Zimmermann, Thomas
; TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)
; FILE REFERENCE: 3,3,3-TRIFLUORO-2-HYDROXY-2-METHYLPROPIONIC ACID
; CURRENT APPLICATION NUMBER: US/10/086,082
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US/09/214,679
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Klebsiella oxytoca
US-10-086-082-11
Query Match
Best Local Similarity 6.7%; Score 117; DB 9; Length 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 42 EPHYTIGPYSTPVLTIHFCDR 52
DB 1 EPHYTIGPYSTPVLTIHFCDR 21
RESULT 5
US-10-004-551-18
; Sequence 18, Application: US/10004551
```





; CURRENT APPLICATION NUMBER: US/10/004,551  
; CURRENT FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 09/635,949  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 110  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 526  
; TYPE: PR1  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: n 2093 can be A, G, C, or T.  
US-10-004-551-24

Query Match 5.8%; Score 101; DB 9; Length 526;

Best Local Similarity 23.9%; Pred. No. 0.65;  
Matches 60; Conservative 28; Mismatches 105; Indels 58; Gaps 11;  
QY 36 TIGPSTPVLITTEPGDRIVTDRAFEAGINSEQIPSOILKMPFLNPGNGPIKVNCAEK 95  
DB 26 TVKGAGPGIETDGLTAAPTPEQPERGVHVTAPT--LKLNLHHPLEEFLEQEGLEK 83  
QY 96 GDVLAVYIESMLPRGVDVPGICAMIPHEGGLGTDLTAMLNDDPLPEKVRMKLDSEKVVW 155  
DB 84 GD---EELRPALPQFDP-----PAPFTPSPLP---RLANQDSREVFT 120  
QY 156 SKRHTL---PYKPHICTLSVSPIDS--INSLIPDNHGNNMDVDPDIPGSGI--TYPLVRA 208  
DB 121 SPTPMAAAVPTQOSKEGPMSPESPMRLITAPLPGPSMAVPTLGPGEIASTTPPSRA 180  
QY 209 -----PG--GRLEFGD-----ANACGGDGE--LCGTAVEFAS--TTIKV 242  
DB 243 DLIKNWQLSWP 253  
QY 241 PGPCSNFSGP 251

## RESULT 9

US-10-004-551-14  
; Sequence 14, Application US/10004551  
; Publication No. US20030004310A1  
; GENERAL INFORMATION:  
; APPLICANT: SHIMKETS, RICHARD A  
; APPLICANT: FERNANDES, ELMA  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY  
; FILE REFERENCE: 15966-559  
; CURRENT APPLICATION NUMBER: US/10/004,551  
; CURRENT FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 09/635,949  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 110  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 993  
; TYPE: PR1  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: n 1755 can be A, G, C, or T  
US-10-004-551-14

Query Match 5.8%; Score 101; DB 9; Length 993;

Best Local Similarity 23.9%; Pred. No. 0.65;  
Matches 60; Conservative 28; Mismatches 105; Indels 58; Gaps 11;  
QY 36 TIGPSTPVLITTEPGDRIVTDRAFEAGINSEQIPSOILKMPFLNPGNGPIKVNCAEK 95  
DB 26 TVKGAGPGIETDGLTAAPTPEQPERGVHVTAPT--LKLNLHHPLEEFLEQEGLEK 83  
QY 96 GDVLAVYIESMLPRGVDVPGICAMIPHEGGLGTDLTAMLNDDPLPEKVRMKLDSEKVVW 155  
DB 84 GD---EELRPALPQFDP-----PAPFTPSPLP---RLANQDSREVFT 120  
QY 156 SKRHTL---PYKPHICTLSVSPIDS--INSLIPDNHGNNMDVDPDIPGSGI--TYPLVRA 208

DB 121 SPTPMAAAVPTQOSKEGPMSPESPMRLITAPLPGPSMAVPTLGPGEIASTTPPSRA 180  
QY 209 -----PG--GRLEFGD-----ANACGGDGE--LCGTAVEFAS--TTIKV 242  
DB 243 DLIKNWQLSWP 253  
QY 241 PGPCSNFSGP 251  
RESULT 10  
US-10-004-551-16  
; Sequence 16, Application US/10004551  
; Publication No. US20030004310A1  
; GENERAL INFORMATION:  
; APPLICANT: SHIMKETS, RICHARD A  
; APPLICANT: FERNANDES, ELMA  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY  
; FILE REFERENCE: 15966-559  
; CURRENT APPLICATION NUMBER: US/10/004,551  
; CURRENT FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 09/635,949  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 110  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 994  
; TYPE: PR1  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: n 1755 can be A, G, C, or T.  
US-10-004-551-16

Query Match 5.8%; Score 101; DB 9; Length 994;

Best Local Similarity 23.9%; Pred. No. 0.65;  
Matches 60; Conservative 28; Mismatches 105; Indels 58; Gaps 11;  
QY 36 TIGPSTPVLITTEPGDRIVTDRAFEAGINSEQIPSOILKMPFLNPGNGPIKVNCAEK 95  
DB 26 TVKGAGPGIETDGLTAAPTPEQPERGVHVTAPT--LKLNLHHPLEEFLEQEGLEK 83  
QY 96 GDVLAVYIESMLPRGVDVPGICAMIPHEGGLGTDLTAMLNDDPLPEKVRMKLDSEKVVW 155  
DB 84 GD---EELRPALPQFDP-----PAPFTPSPLP---RLANQDSREVFT 120  
QY 156 SKRHTL---PYKPHICTLSVSPIDS--INSLIPDNHGNNMDVDPDIPGSGI--TYPLVRA 208  
DB 121 SPTPMAAAVPTQOSKEGPMSPESPMRLITAPLPGPSMAVPTLGPGEIASTTPPSRA 180  
QY 209 -----PG--GRLEFGD-----ANACGGDGE--LCGTAVEFAS--TTIKV 242  
DB 243 DLIKNWQLSWP 253  
QY 241 PGPCSNFSGP 251

## RESULT 11

US-10-004-551-26  
; Sequence 26, Application US/10004551  
; Publication No. US20030004310A1  
; GENERAL INFORMATION:  
; APPLICANT: SHIMKETS, RICHARD A  
; APPLICANT: FERNANDES, ELMA  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY  
; FILE REFERENCE: 15966-559  
; CURRENT APPLICATION NUMBER: US/10/004,551  
; CURRENT FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 09/635,949  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 110





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 15:42:52 ; Search time 41 seconds  
(without alignments)  
769,075 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1748

Sequence: 1 MKWLEESIMAKRGVAGRKPK.....GNMVPKVTGAMLNKNIIV 329

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 284224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:\*\*

1: Pirl:\*\*

2: Pirl:\*\*

3: Pirl:\*\*

4: Pirl:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	363	20.8	300	2 A83653	acetamidase BH0025
2	342	19.6	338	2 B90458	hypothetical prote
3	332.5	19.0	377	2 C72750	probable ace-amida
4	322.5	18.4	304	2 H87572	amidase-related pr
5	304.5	17.4	298	2 E75139	hypothetical prote
6	300.5	17.2	285	2 G72414	hypothetical prote
7	269	15.4	328	2 A81894	hypothetical prote
8	261.5	15.0	259	2 A81776	hypothetical prote
9	233.5	13.4	410	2 I39115	formamidase-like p
10	233	13.3	296	2 A21400	hypothetical prote
11	208	11.9	432	2 F04712	probable formamida
12	204.5	11.7	407	2 S74213	formamidase (EC 3.
13	179	10.2	389	2 E71097	hypothetical prote
14	178.5	10.2	406	2 A47596	acetamidase - Myco
15	130	7.4	307	2 T04713	probable formamida
16	107	6.1	238	2 A81298	hypothetical prote
17	102	5.8	238	2 A81670	hypothetical prote
18	100.5	5.7	617	2 T49535	probable multifunc
19	98.5	5.6	948	2 B81883	exonuclease ABC c
20	97.5	5.6	504	2 S74034	amidase (EC 3.5.1.
21	97.5	5.6	949	2 A81138	exonuclease ABC c
22	96.5	5.5	432	2 H96927	inactivated probab
23	94.5	5.4	470	2 A81861	hypothetical prote
24	94	5.4	655	2 G75582	cell division prot
25	94	5.4	857	1 JQ1525	nitrate reductase
26	92.5	5.3	505	2 G75197	lacZ expression re
27	92	5.3	501	1 A40938	cytochrome P450 ib
28	92	5.3	822	2 C86431	T518.5 protein - A
29	90.5	5.2	500	2 C69580	alpha-L-arabinofur

30 90 5.1 49: 2 E93356  
31 89.5 5.1 688 2 A44306  
32 89.5 5.1 698 2 C91049  
33 89.5 5.1 688 2 G85893  
34 89 5.1 485 2 C69584  
35 89 5.1 704 2 G83950  
36 88.5 5.1 428 2 G72037  
37 88.5 5.1 428 2 B85586  
38 88.5 5.1 467 2 B97213  
39 88 5.0 773 2 G86856  
40 88 5.0 868 1 JC4283  
41 88 5.0 1378 2 G88637  
42 88 5.0 1849 2 C41859  
43 87.5 5.0 652 2 AG0031  
44 87.5 5.0 1661 2 A83695  
45 86.5 4.9 376 2 A45634

#### ALIGNMENTS

RESULT 1

A83653

acetamidase BH0025 [imported] - Bacillus haedurans (strain C-125)

C:Species: Bacillus haedurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: A83653

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;

Nucleic Acids Res. 28, 4317-4331, 2000

A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: A83653

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-300 <STO>

A:Cross-references: G3:AF001507; G3:BA000004; NID:g1c172612; PIDN:BA803744.1; CSPD3:

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0025

Query Match 20.8%; Score 363; DB 2; Length 300;  
Best Local Similarity 32.4%; Pred. No. 1.1e-22;  
Matches 94; Conservative 47; Mismatches 127; Indels 22; Gaps 6;

QY 29 MQSEFHTTIGPSTPVLTTEPGRRIIVDTREAFEGAI-NSEQDIPSQLLKPPFLNPQNGP 87  
DB 13 MDKE-----NVFIASCSGDTVFEIKDCGFSQITNEQALTS--IDFNVRNPATGP 62  
QY 88 IYNGAEKGDVLAIVTESMLPRGVDPYG-CAMIPHFGGLTGDTLTAMLNDPLPEKVRMK 147  
DB 63 LYVEGARRGDMLEIE---LDIKVGKGVTAPAGLGALES-----LNSP---TTKLFP 111  
QY 148 LSEKVVNKRHFLPYKPHIGTLSVSPEDSINSITPDNHHGNMVDVDPGGSIIYPLVR 207  
DB 112 IEGDDVVYTGRLPLQFMIGVIGTAPGGERINNGTTPGHGNGEDTKQIKGTTVYLVE 171  
QY 208 APGGRIFGDHACQGDGEICGTAVEFASITIKVDLIKMWLSWPRNAENIMSIGSA 267  
DB 172 VOCALLALGDLHAMGCGEILICGVEIAGTVTLKVNKKERMFPLPALKTTHFWMTIASA 231  
QY 268 RPLFDATRIAYRDLIYLVVEDEFGQNDAYMLLSQCGKVRKGNMVDPKYT 317  
DB 232 ETIDAAVAOATKNMATFLANRTALSTSEAGMLLSCAGDLYVSQIVNPLKT 281

RESULT 2

B90458

hypothetical protein SS02810 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C:Accession: B90458

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awayer, M.J.; C  
Jong, J.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redde

```

156 SKRH---TLPYKPHIGIT-SVSP-----EI-----176
      : : : : : : : : : : : : : : : : : : : :
131 FSEDIQNRVTFGDPFLGVMGAPSKELLAEIKERDRLLKRGFVLPTPEGAVPPPEFV 190
      : : : : : : : : : : : : : : : : : : : :
177 --DS--NSLTEDNHGKNDVPDIOGSGIITPLVRAFGRLFIQSAHACQGGRETCGTAVEF 234
      : : : : : : : : : : : : : : : : : : : :
191 ASGLRTTFPENGKGLDVHVFSPGSKIYPPFVEGALFSVGDHAHYAQOIGECVGTAIEM 250
      : : : : : : : : : : : : : : : : : : : :
235 ASITTTIKVDLI---KNMOLSWPRMENAEN-----IMSIGSARPLEDATR 275
      : : : : : : : : : : : : : : : : : : : :
251 GAVATLRFGVISGAKKYGCFPIEPKPSSESVRRFTSHKHVATGVGVIDSLASENATL 310
      : : : : : : : : : : : : : : : : : : : :
275 IATKDLIYV---JVEDREFEWDNAYMQLSQCKQVRLGNMVD-PRYTYGAMLNKNLLV 328
      : : : : : : : : : : : : : : : : : : : :
311 SAKPALLNLINLEKAGYTRFQAYELASVAADRLLSQLDVDPNFTVTAFLPLDIFI 366
      : : : : : : : : : : : : : : : : : : : :

RESULT 4
#87572
a:case-related protein [imported] - Caulobacter crescentus
c:Species: Caulobacter crescentus
c:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
c:Accession: H87572
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.I.; Nelson, K.E.; Eisen, J.; Heidelberg,
P.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K
C.; Errolla, A.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB87249; MUID:21173698; PMID:11259647
A:Accession: H87572
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <STO>
A:CROSS-references: GB:AE005673; NID:gl3424186; PIDN:AAK24580.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2611

Query Match 18.4%; Score 322.5; DB 2; Length 304;
Best local Similarity 31.7%; Pred. No. 2.8e-19;
Matches 99; Conservative 41; Mismatches 129; Indels 43; Gaps 11;

QY 43 PVLITIEGDRLLVDTKDAFECAI-----NSEQDIQSLLIKMPELNPQ-----NGPTM 89
      : : : : : : : : : : : : : : : : : : : :
DB 8 PVLITVNSGDSVTV---ECVSGGVFVMPPEASPYAIPSAQIAIHDSNPRRGPHILITGPIA 64
      : : : : : : : : : : : : : : : : : : : :
QY 90 VNGAEKGDVLAVIESMLPRGVDPYGTICAMIPFGGLTGTDLTAMLNDPLPEK-VRMIKL 148
      : : : : : : : : : : : : : : : : : : : :
DB 65 IAGAEPCDLEUVRIEA-EPN-NDWGVCAVP---LAGI-----LPEDFPERVSVSHAV 113
      : : : : : : : : : : : : : : : : : : : :
QY 149 DSE-----KVYWSRRHILPKPHIGTILSVSP--FIDSINSLTPNHGGMNDVPDIOGPGSIT 202
      : : : : : : : : : : : : : : : : : : : :
DB 124 DAORGVCCKPEWGPE--LPLAFFGIMGVAPAKYKGLSREPREHGNMKNKLAVAGSTL 171
      : : : : : : : : : : : : : : : : : : : :
QY 203 YPLVRAPGGRI:FGIDAHACQGGREICGTAVEFASITTIKVDLIKNQ-----LSWPRM 255
      : : : : : : : : : : : : : : : : : : : :
DB 172 YLPVWVPGANFSYGDGHGCGDCEVCVNALENGLTGTFTFVLHKKANGAADIAPFAWPA 231
      : : : : : : : : : : : : : : : : : : : :
QY 256 ENENEMTSIGSARPLEDATRIATYRLIYWLVDVFGEQNDAYMQLSQCKQVRLGNMVDPK 315
      : : : : : : : : : : : : : : : : : : : :
DB 232 ETETHYVLMGFHREDLDMKQALROTIDFITARSNLRTVQAYGCSLAVDFRVIQTWNGE 251
      : : : : : : : : : : : : : : : : : : : :
QY 315 YIVGAMLNKNLL 327
      : : : : : : : : : : : : : : : : : : : :
DB 292 KGVHALLSKGLL 303
      : : : : : : : : : : : : : : : : : : : :

RESULT 5
#E75139
a:cytotoxic protein PA0614 - Pyrococcus abyssi (strain Orsay)
c:Species: Pyrococcus abyssi
c:Date: 23-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
c:Accession: E75139
R:Anonymous. Genoscope

```

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: F75139  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-298 <RAW>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:g545867; PIDN:CA849830.1; PID:g545834  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0614

Query Match 17.4%; Score 304.5; DB 2; Length 298;

Best Local Similarity 30.7%; Pred. No. 9.7e-18;

Matches 94; Conservative 45; Mismatches 120; Indels 47; Gaps 10;

Qy 35 YTGPGYSTPVLTPGDRIVTDRAFGAINEQDIPSOILKMPF--LNPNQGPIMVNG 92  
Db 14 YSGPNNKEVARAKPGEVVFTDLALGGQVASED---TIEKIDFSRVNATGPLYVEG 70  
Qy 93 AEKGDVLAVYIESMLPRGYDPYIGICAMIPHFGLTGDITAMNDPLPEKV-----RMK 147  
Db 71 AKRGGLRVLDLIDIKVEG-----KGAVVTAPGAGVGGKKVEEPOTRICE 114  
Qy 148 DSEKVVYKSRHTLPYKPHIGTISVSPETLSINSLTPDNGNMDVFDIGSGSTYPPYR 207  
Db 115 VKDGFVIF-KGKIPAMPILGV-GVAYD-EEVPTDTPKSHGGNMDINIKQTTFYFPVP 172  
Qy 208 APOGRFEGDAHACOGDGTGCTGTAFFASITTKVDLIKMWLSWPMNAENIMSGAR 267  
Db 173 VDCAVTAIGDLHACVAGGVCVACVSGEVVTVTPMG-KLEAFLETENSTYVAVSD 211  
Qy 268 RPLEDATRIAYRDLIYWLVEDCFGE-----QMD-AYML-SOCCGVRLGNMVPKVTVCAMLNK 324  
Db 232 ENLDKALEAV-----SIGVEALRKSNDLSMDYAMGLASIVHDLVLSQNDPRKIV 292  
Qy 319 GAMLNK 324  
Db 283 NVRIPK 288

RESULT 6

G72414

A:Title: Hypothetical protein TM0119 - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jun-2000

C:Accession: G72414  
R:Neeson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickox Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and bacteria from genome sequencing of *Thermotoga maritima*

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: G72414

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-285 <ARN>

A:Cross-references: GB:AF001598; GB:AE000512; NID:g4980509; PIDN:AAD35213.1; PID:g498061

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0119

Query Match 17.2%; Score 300.5; DB 2; Length 285;

Best Local Similarity 28.7%; Pred. No. 1.8e-17;

Matches 85; Conservative 53; Mismatches 125; Indels 33; Gaps 9;

Qy 35 YTGPGYSTPVLTPGDRIVTDRAFGAINEQDIPSOILKMPF--LNPNQGPIMVNG 92  
Db 11 YFSANMAPVEEYVPGGVFTDLALGGSYD-----KIDFSKVNATGPGVFNAG 60  
Qy 93 AEKGDVLAVYIESMLPRGYDPYIGICAMIPHFGLTGDITAMNDPLPEKVYKMKLQSE 151  
Db 61 VKPGDTLKVRIKRIELPR-----RGMIVTKGFG-----VIGSEV-EGFTHFLEIE 104

Qy 152 KVVSKRRH---TLPPYKPHIGTISVSPEDINSILPNDHGGNMDVPDIPGSGITYPLVRA 208  
Db 107 K--WAVLEFGVRIPIHPMYGVGVGVAQEGEYPTGTAHRRGGNMDTKEITENVVHLPVFQ 164  
Qy 209 PGGRIPIGDAHACOGDGTGCTGTAFFASITTKVDLIKMWLSWPMNAENIMSGAR 268  
Db 155 EGALLALGDVHATMGDGCVCVACVSGEVVTVTPMG-KLEAFLETENSTYVAVSLP 223  
Qy 269 PLEDATRIAYRDLIYWLVEDCFGEQMDAYML-SOCCGVRLGNMVPKVTVCAMLNK 324  
Db 224 DIEALKVETRETVMFIQRRKTIPTFDAYMLASLSVDVGI SOLVNPAPKTAKARIKP 279

RESULT 7

AH1894

A:Title: Hypothetical protein all0706 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AH1894

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriig Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Iabat DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Nostoc sp.* strain PCC 7120

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AH1894

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-328 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA872663.1; PID:g17130051; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all0706

Query Match 15.4%; Score 269; DB 2; Length 328;

Best Local Similarity 27.5%; Pred. No. 9.1e-15;

Matches 91; Conservative 59; Mismatches 140; Indels 40; Gaps 9;

Qy 21 VTHILTEEMOKERHYTIGPYS---TPVLTEPDGRIIVTDRAFGAINEQDIPSOILK 77

Db 1 MTHILKATKSVH--DGGFSLLEFALKVDSSDTIDVETTYG----YIDKAPERT 54

Qy 78 MPFLN-----PQ---NGPIMVNGAKGKGVLAIVYFESMLPRGVDPVGCAMLP 121

Db 55 PFDLDICNLPPEKKTANGPHELTGTIIYVKDAPGDLVLEKLEAIAPS--LPVGFNAIRS 112

Qy 122 HFGGLTGDLTAMLNDPLPEKVRMKL---DSEKVVYKSRHTLPYKPHIGTISVSPETD 177

Db 113 GWG-----ALPNOFTQALRFILNLANNTAEFPNPSGKIKIPLTPFFGLGVAATPEN 164

Qy 178 SINSLTEPDHGGNMDVPDIPGSGITPVLVAPGGRIFIGDAHACOGDGTGCTGTAFFASI 237

Db 165 ARSSVPVPGYGGNLDNRELQAGSRIFLPFVPGCLSLDGGHSAQDGGVNNVTAIETSMN 224

Qy 238 TTKVDLIKMWLSWPMNAENIMSGARPLEDATRIAYRDLIYWLVEDCFGEQMDAY 297

Db 225 GRIQLTLRKDLHFTAPIAETPTDIITMGFAPTLDAALEQALKNMIDFLERFVNLSPEY 284

Qy 298 MLLSOCCGVRLGNMVD-PKYTVYCAMLNK 326

Db 285 VLCSLVNFRITQVNVSPNKGVRHGLPKKL 314

RESULT 8

AB1776

A:Title: Hypothetical protein lin2752 [imported] - Listeria innocua (strain Clip11262)

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AB1776

R:Glaser, P.; Brangeul, L.; Buchricser, C.; Amed, A.; Baquero, F.; Berche, P.; Bloch

G.; Dominguez-Serna, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi

G.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001  
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M  
 ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
 A:Title: Comparative genomics of *Listeria* species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AB1776  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-295 <GLA>  
 A:Cross-references: GB:AI592022; PIDN:CA097378.1; PID:gl6412098; GSPDB:GN00178  
 A:Experimental source: strain Clp11262  
 C:Genetics:  
 A:Gene: lin2752

Query Match 15.0%; Score 261.5; DB 2: Length 299;  
 Best Local Similarity 29.0%; Pred. No. 3,4e-14;  
 Matches 79; Conservative 44; Mismatches 112; Indels 37; Gaps 7;  
 QY 43 PVLTIIEPGDRIIVDTTRCAFEGAINSEODIPSOILKMPFLNPQNGPIWNGAEKSDVLAAY 122  
 DB 20 PVLTVXDSGVVITIKIDHFGQINKELHYGEIDWKQF-SPITGPTICIEARPGMLAVT 78  
 QY 103 IESMLPRGVDVPGICAMIPHFGGLTGDTLAMLNCP-----LPEK-VRMTKLUSEKV 153  
 DB 79 IEXIT-----LISKDV-VILNGHTIGVIDDOLLNNCIRKIKQNK: 119  
 QY 154 YWKRHTLPYKPHIGTSLVSPEDISNLS-----TPDNHGNMVDVPIGPGSITYPIVLRAP 209  
 DB 120 IYSDE-HVQLQKTIQL-----KTES-HOTKPPHMTKKGGLDSDPOLTECATIFLPEVKF 175  
 QY 210 GGRFLIGDAHACOGDEICGTAVEFASITIKVDLKNMOLSPRMEANENISGSASP 269  
 DB 176 GAMLHVADIRATGFGKTTSTSAEVAEVTIRLQLLNKNTAPTPIIHKNNLICLGSALT 235  
 QY 270 LEDATRIAYDLIYLVEDFGFQWDAYMLLS 301  
 DB 236 TERATKALQNLNVLMSDKITLEDAPILLS 267

RESULT 9  
 T39115  
 formamide-like protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T39115  
 R:Runt, C.; Aves, S.; Mcbougall, R.C.; Rajandream, M.A.; Barrell, H.G.  
 Submitted to the EMBL Data Library, November 1999  
 A:Reference number: Z21829  
 A:Accession: T39115  
 A:Status: preliminary; translated from GB/EMBL/CDHUJ  
 A:Molecule type: DNA  
 A:Residues: 1-410 <HUN>  
 A:Cross-references: EMBL:AL132779; PIDN:CA060014.1; GSPDB:GN00065; SPDB:SPAC869.04  
 A:Experimental source: strain 972h-; cosmid c869  
 C:Genetics:  
 A:Gene: SPDB:SPAC869.04  
 A:Map position: 1

Query Match 13.4%; Score 233.5; DB 2: Length 410;  
 Best Local Similarity 26.1%; Pred. No. 1.1e-11;  
 Matches 92; Conservative 42; Mismatches 115; Indels 57; Gaps 17;  
 QY 46 TIFPGDRIIVDTTRAFGA-----NSEQDIPS-QLLXMPFLNTONGPIWNGAEKSDVLAAY 101  
 DB 34 SIDGETIVKTECLDWTGGC-KNDSAEITKNVDLTKIHYL---SGPFEKGAEPGSDVWV 90  
 QY 102 YIESMLPRGVDVPGICAM:PHFGGLTGDTLAMLNCPILPEKVRMIKLDSEKVVNSKSH-- 159  
 DB 91 EIQDVPLENQPGVSGSIFAKENG-----GGFLDEHYKAAKAV-FDEG-FCSSRHLP 143  
 QY 160 -----TLPIK-----PR-----IGTSS-- 171  
 DB 144 GVRFPGLIHPGLIGTAPSKELAEWRNRREGVAENPHSTHYWAOLEFNASYAFAGLAE 203

QY 172 -VSPEDISINSLT-----PUNHGNMVDVPIGPGSITYPIVLRAPGGKLFIGDAHACQGDGE 226  
 DB 204 KESATVASEGARTIFGRPEN-GGNCIDKNLSRGSKVFLPVHVPGAKLSIGDLHFSGODGE 262  
 QY 227 1--CGTAVEFAS--TIKVVLKINW-----QLSWPRMEANENISIGS 266  
 DB 263 ISFCG-AIEAGSITIKKILKNG:SDLAMKSPYLPQGVPEHFFSFSRYLTIFEGESVDES 321  
 QY 267 APRLEDAIRIAYRDLIYVVE---DFGFGWDAYMLLSQCKVR--LGNMVD 313  
 DB 322 GKQHYLCITTAIKQCLRIYIEYFRREGYNDYQLYLLLS-CAPIQGHVAGIVD 372

RESULT 10  
 AC1400  
 hypothetical protein lmo2603 [imported] - *Listeria monocytogenes* (strain EGD-e)  
 C:Species: *Listeria monocytogenes*  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC1400  
 K:Glaser, P.; Frangoul, L.; Huchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloes  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
 J.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.  
 ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
 A:Title: Comparative genomics of *Listeria* species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AC1400  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-296 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CA06081.1; PID:gl6412091; GSPDB:GN00177  
 A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo2603

Query Match 13.3%; Score 233; DB 2: Length 296;  
 Best Local Similarity 25.0%; Pred. No. 8e-12;  
 Matches 74; Conservative 62; Mismatches 136; Indels 24; Gaps 7;  
 QY 28 EMKEFEHTYIGPYSTPVLTIIEPGDRIIVDTTRCAFEGAINSEODIPSOILKMPFLNPQNGP 87  
 DB 13 EMKES-----TEPARVKDGVVVKTKTDHFGQIHAKOLHYGEIDWKQF-SPTTGP 63  
 QY 88 INVNGAEKSDVLAAYIESMLPRGVDVPGICAMIPHFGGLTGDTLAMLNCPILPEKVRMIK 147  
 DB 64 IYIEEARPGDLAITIKIELLGTETVFLNG--PNI-GITDILLTS-----NSTRCYK 113  
 QY 148 LDSEKVVNSKSHITLPYKPHIGTSLVSPEDISIN-SLTPDNHGNMVDVPIGPGSITYPIV 206  
 DB 114 VENNOITYSEDIHPIRKTIGLL----KTEELNPSKVPTKNGGLDSDKITEGATIFLPV 169  
 QY 207 RAPGGLRFLIGDAHACOGDEICGTAVEFASITIKVDLKNMOLSPRMEANENISIGS 266  
 DB 170 EYXGASLVHGNVRATIGFGKITATSAEAPAEVTLQIHKNTAPTPIIHYHNLICLAS 229  
 QY 267 APRLEDAIRIAYRDLIYVVEFGFQWDAYMLLSQCKVRGNMVDPKYTVGMAL 322  
 DB 230 GTTIEKATOKTHNNITLITLTSKMTITIEDALFLISLQADFQVCKLCKPNITTSIKL 285

RESULT 11  
 T04712  
 probable formamide (BC 3.5.1.49) F19F18.40 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
 C:Accession: T04712  
 R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K  
 submitted to the Protein Sequence Database, March 1999  
 A:Reference number: Z15382  
 A:Accession: T04712  
 A:Molecule type: DNA

A:Residues: 1-432 <BE>  
A:Cross-references: EMBL:AL035605  
A:Experimental source: cultivar Columbia; BAC clone F19F18  
C:Genetics:  
A:Map position: 4  
A:Intons: 73/3; 153/1; 196/3; 275/2; 356/3; 392/3  
A:Note: F19F18.40  
C:Keywords: hydrolase

Query Match 11.9%; Score 208; DB 2; Length 432;  
Best Local Similarity 24.1%; Pred. No. 1.6e-09;  
Matches 92; Conservative 44; Mismatches 140; Indels 106; Gaps 16;

Dy 28 EMQEFHYTIGYSTPVLITISPGDRIIVCTDAFEAGI-----NSEQDIPQLKMPFLNPQ 84  
Db 20 QODQPLHARHPEIPPAEAKVAGEFFRVEDMAGGVIKNDSDSDIKN--LVGITTHL 77  
Qy 85 NGPIWV-----NGAFKGDVLAIVTESLPRGVDPYGCAMLPHEGGLTGTDLTAMINDPL 139  
Db 78 SGPIRVVDEGVAANKAGDALLAVEICNLGPIPGSEMGFTGSEURENG-----SGFLTDF 141  
Qy 140 PEKVRMKLDSEKVVYKSRHTPEYKP-----HTGLVSP----- 174  
Db 132 P-----CATKAIWFFGSIYAYSPQIGVRFPGTHFSVGTAFSPNLLRNNDRERQ 183  
Qy 175 -EIDSINSL-----TPQ-----NHGGMKVPDVGSGSIYPIVAPAPG 210  
Db 144 LEESVSESLTCEIEETGTPWERIANZAARTIPGRNGGNCIDIKLSRGSKYLPVFEV 243  
Qy 211 GRLEFGDAHACOGGGEI--CGTAVEFASITTIKVDLIKNNOLSW-----PRMENAEINMS 263  
Db 244 ANLSTGDMHFSGQGEISFGC-AIEMSGPELKEIIRNGMORYLTPMGTPPLHVNPIPE 302  
Qy 264 IGSARE-----LEDAIRIAYSDI IYWLVE---DRGPFQKAYWEL 300  
Db 303 IGPVFRFSENVJFEGISVSGKQHYLDAT-VAYKAVLUNA:DYLFKFGYSKEQVY:LL 362  
Qy 301 SQCG-KVRLGNMVPKVTVAAM 321  
Db 362 SCCPCGR:SGIVDSPNAVATL 383

RESULT 12  
S74213  
formamide (EC 3.5.1.49) A - Methylophilus methylotrophus  
C:Species: Methylophilus methylotrophus  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: S74213; S78115  
R:Wyborn, N.R.; Mills, J.; Williams, S.G.; Jones, C.W.  
Eur. J. Biochem. 240, 314-322, 1996  
A:Title: Molecular characterisation of formamide from Methylophilus methylotrophus.  
A:Reference number: S74213; MUID:96439058; PMID:8841393  
A:Accession: S74213  
A:Molecule type: DNA  
A:Residues: 1-407 <W>  
A:Cross-references: EMBL:X96632; NID:q1480104; PIDN:CAA67953.1; PID:q1480105  
A:Experimental source: strain NCIB 10515  
A:Accession: S78115  
A:Molecule type: protein  
A:Residues: 1-544-47;172-176;211-215 <W>  
C:Genetics:  
A:Gene: fmdA  
C:Keywords: hydrolase

Query Match 11.7%; Score 204.5; DB 2; Length 407;  
Best Local Similarity 25.5%; Pred. No. 3e-09;  
Matches 84; Conservative 38; Mismatches 116; Indels 91; Gaps 14;

Dy 28 EMQEFHYTIGYSTPVLITISPGDRIIVCTDAFEAGI-----NSEQDIPQLKMPFL 81  
Db 15 EQQGLHNHRHDPJPMIAWPKGDFRVECDMTGCGGNDSANDVRDV--DLQVHVE 72  
Qy 82 NPONGPIWNGAEKGDVLAIVTESLPRGVDPYGCAMLPHEGGLTGTDLTAMINDPL 140

A:Residues: 1-173 <BE>  
A:Cross-references: EMBL:AL035605  
A:Experimental source: cultivar Columbia; BAC clone F19F18  
C:Genetics:  
A:Map position: 4  
A:Intons: 73/3; 153/1; 196/3; 275/2; 356/3; 392/3  
A:Note: F19F18.40  
C:Keywords: hydrolase

Query Match 11.9%; Score 208; DB 2; Length 432;  
Best Local Similarity 24.1%; Pred. No. 1.6e-09;  
Matches 92; Conservative 44; Mismatches 140; Indels 106; Gaps 16;

Dy 28 EMQEFHYTIGYSTPVLITISPGDRIIVCTDAFEAGI-----NSEQDIPQLKMPFLNPQ 84  
Db 20 QODQPLHARHPEIPPAEAKVAGEFFRVEDMAGGVIKNDSDSDIKN--LVGITTHL 77  
Qy 85 NGPIWV-----NGAFKGDVLAIVTESLPRGVDPYGCAMLPHEGGLTGTDLTAMINDPL 139  
Db 78 SGPIRVVDEGVAANKAGDALLAVEICNLGPIPGSEMGFTGSEURENG-----SGFLTDF 141  
Qy 140 PEKVRMKLDSEKVVYKSRHTPEYKP-----HTGLVSP----- 174  
Db 132 P-----CATKAIWFFGSIYAYSPQIGVRFPGTHFSVGTAFSPNLLRNNDRERQ 183  
Qy 175 -EIDSINSL-----TPQ-----NHGGMKVPDVGSGSIYPIVAPAPG 210  
Db 144 LEESVSESLTCEIEETGTPWERIANZAARTIPGRNGGNCIDIKLSRGSKYLPVFEV 243  
Qy 211 GRLEFGDAHACOGGGEI--CGTAVEFASITTIKVDLIKNNOLSW-----PRMENAEINMS 263  
Db 244 ANLSTGDMHFSGQGEISFGC-AIEMSGPELKEIIRNGMORYLTPMGTPPLHVNPIPE 302  
Qy 264 IGSARE-----LEDAIRIAYSDI IYWLVE---DRGPFQKAYWEL 300  
Db 303 IGPVFRFSENVJFEGISVSGKQHYLDAT-VAYKAVLUNA:DYLFKFGYSKEQVY:LL 362  
Qy 301 SQCG-KVRLGNMVPKVTVAAM 321  
Db 362 SCCPCGR:SGIVDSPNAVATL 383

RESULT 13  
E71097  
hypothetical protein PH1041 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 18-Aug-2000  
C:Accession: E71097  
R:Kawarabayashi, Y.; Swada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; O  
X.; Onikubo, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; O  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil  
A:Reference number: A71000; MUID:98344137; PMID:9679134  
A:Accession: E71097  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-389 <NA>  
A:Cross-references: GB:AP000004; NID:93236131; PIDN:BA430139.1; PID:dl031082; PID:9  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by Gen  
C:Genetics:  
A:Gene: PH1041  
C:Superfamily: pyrococcus horikoshii hypothetical protein PH1041

Query Match 10.2%; Score 179; DB 2; Length 389;  
Best Local Similarity 22.5%; Pred. No. 3.7e-07;  
Matches 69; Conservative 32; Mismatches 86; Indels 120; Gaps 11;

Dy 88 IMVNGAEKGVLAIVTESN-----LPRGVDPYGCAMLPHEGGLTGTDLTA-----MLND 137  
Db 16 VAVEGAKVGDATATIRVERIRVLSKAASGVDS-----PREGSVGDPYVAKKCPVNE 68  
Qy 138 PIPE-----KVRMTK-----LDSE-----KVW 155  
Db 69 PWPEFEVWIGEDAIRCKKCGAPVSPFRMNGYTWVFDHEGGVGVTVYDKKIAEMIAKNW 128  
Qy 156 SKRHITLP-----YKPHIGTISVSPETSDINSLSLPDWHG----- 188  
Db 129 -EWHALPKNSKOVPIILYAKADIVCPSPRIKPIGQLGTIPAVD-----IPDSHNGDFG 182  
Qy 189 -----GNMVDPIGPGSIYPLVRAPGGRGLFTGDHACOGDG 225  
Db 183 YFLINAPHPYAIKEDYETKLTGDHLDVDSVREGAIVIAVPKVEGGVYAGDAHAHEGDD 242  
Qy 226 EICGTAVEFASITTIKVDLIKNNOLSWPRMENAEINMSISGARPLEDATRIAYFDLIYWL 285  
Db 243 EVAGHTTDVTAETIVTEVIKGLNLEGPILLPEEDLP-----PLAKPWKRDEWEKVERL 297  
Qy 286 VEDFGPE 292  
Db 298 AKKEGIE 304

RESULT 14  
A47696

A:Residues: 1-432 <BE>  
A:Cross-references: EMBL:AL035605  
A:Experimental source: cultivar Columbia; BAC clone F19F18  
C:Genetics:  
A:Map position: 4  
A:Int-ONS: 73/3; 153/1; 196/3; 275/2; 356/3; 392/3  
A:Note: F19F18.40  
C:Keywords: hydrolase

Query Match 11.9%; Score 208; DB 2; Length 432;  
Best Local Similarity 24.1%; Pred. No. 1.6e-09;  
Matches 92; Conservative 44; Mismatches 140; Indels 106; Gaps 16;

Dy 28 EMQKEFHYTIGYSTPVLITISPGDRIIVCTDAFEAGI-----NSEQDIPQLKMPFLNPO 84  
Dy 20 QODQPLHARHPEIPPAEAKVAGEFFRVEDMAGGVIKDNDASDIKN--LVGITTHL 77  
Dy 85 NGPIWV-----NGAFKGDVLAIVTESLPRGVDPYGCAMLPHEGGLTGTDLTAMINDPL 139  
Dy 78 SGPIRVVDEGVAANKAGDALLAVEICNLGPIPGSEMGFTGSEURENG-----SGFLTDF 141  
Dy 140 PEKVRMKLDSEKVVYKSRHTPEYKP-----HTGLSVSP----- 174  
Dy 132 P-----CATKAIWFFGSIYAYSPQIGVRFPGTHPSVGTAFSPNELLKNNDRERQ 183  
Dy 175 -EIDSINSL-----TPQ-----NHGGMKVPDVGSGSIYPIVAPAPG 210  
Dy 144 LEESVSESLTCEIEETGTPWERIANZAARTIPGRNGGNCIDIKLSRGSKYLPVFEV 243  
Dy 211 GRLEFGDAHACOGGGEI--CGTAVEFASITTIKVLDIKNNOLSW-----PRMENAEINMS 263  
Dy 244 ANLSTGDMHFSQGGCEISFGC-AIEMSGPELKECEIRNGMRYLTPTMGTPLHVNPIPE 302  
Dy 264 IGSARE-----LEDAIRIAYSDI IYWLVE---DRGPFQKAYWEL 300  
Dy 303 IGPVFRFSENVJFEGISVSGKQHYLDAT-VAYKAVLUNA:DYLFKFGYSKEQVY:LL 362  
Dy 301 SQCG-KVRLGNMVPKVTVAAM 321  
Dy 362 SCCPCGR:SGIVDSPNAVATL 383

RESULT 12  
S74213  
formamide (EC 3.5.1.49) A - Methylophilus methylotrophus  
C:Species: Methylophilus methylotrophus  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: S74213; S78115  
R:Wyborn, N.R.; Mills, J.; Williams, S.G.; Jones, C.W.  
Eur. J. Biochem. 240, 314-322, 1996  
A:Title: Molecular characterisation of formamide from Methylophilus methylotrophus.  
A:Reference number: S74213; MUID:96439058; PMID:8841393  
A:Accession: S74213  
A:Molecule type: DNA  
A:Residues: 1-407 <W>  
A:Cross-references: EMBL:X96632; NID:q1480104; PIDN:CAA67953.1; PID:q1480105  
A:Experimental source: strain NCIB 10515  
A:Accession: S78115  
A:Molecule type: protein  
A:Residues: 1-544-47; 172-176; 211-215 <W>  
C:Genetics:  
A:Gene: fmdA  
C:Keywords: hydrolase

Query Match 11.7%; Score 204.5; DB 2; Length 407;  
Best Local Similarity 25.5%; Pred. No. 3e-09;  
Matches 84; Conservative 38; Mismatches 116; Indels 91; Gaps 14;

Dy 28 EMQKEFHYTIGYSTPVLITISPGDRIIVCTDAFEAGI-----NSEQDIPQLKMPFL 81  
Dy 15 EQQGLHNHRHDPJPMIAWPKGDFRVECDMTGCGGNDSANDVRDV--DLQVHVE 72  
Dy 82 NPONGPIWNGAEKGDVLAIVTESLPRGVDPYGCAMLPHEGGLTGTDL--FAMLDNF:FP 140

A:Residues: 1-122 <BE>  
A:Cross-references: EMBL:AL035605  
A:Experimental source: cultivar Columbia; BAC clone F19F18  
C:Genetics:  
A:Map position: 4  
A:Int-ONS: 73/3; 153/1; 196/3; 275/2; 356/3; 392/3  
A:Note: F19F18.40  
C:Keywords: hydrolase

Query Match 11.9%; Score 208; DB 2; Length 432;  
Best Local Similarity 24.1%; Pred. No. 1.6e-09;  
Matches 92; Conservative 44; Mismatches 140; Indels 106; Gaps 16;

Dy 28 EMQKEFHYTIGYSTPVLITISPGDRIIVCTDAFEAGI-----NSEQDIPQLKMPFLNPO 84  
Dy 20 QODQPLHARHPEIPPAEAKVAGEFFRVEDMAGGVIKDNDASDIKN--LVGITTHL 77  
Dy 85 NGPIWV-----NGAFKGDVLAIVTESLPRGVDPYGCAMLPHEGGLTGTDLTAMINDPL 139  
Dy 78 SGPIRVVDEGVAANKAGDALLAVEICNLGPIPGSEMGFTGSEURENG-----SGFLTDF 141  
Dy 140 PEKVRMKLDSEKVVYKSRHTPEYKP-----HTGLSVSP----- 174  
Dy 132 P-----CATKAIWFFGSIYAYSPQIGVRFPGTHPSVGTAFSPNELLKNNDRERQ 183  
Dy 175 -EIDSINSL-----TPQ-----NHGGMKVPDVGSGSIYPIVAPAPG 210  
Dy 144 LEESVSESLTCEIEETGTPWERIANZAARTIPGRNGGNCIDIKLSRGSKYLPVFEV 243  
Dy 211 GRLEFGDAHACOGGGEI--CGTAVEFASITTIKVLDIKNNOLSW-----PRMENAEINMS 263  
Dy 244 ANLSTGDMHFSQGGCEISFGC-AIEMSGPELKECEIRNGMRYLTPTMGTPLHVNPIPE 302  
Dy 264 IGSARE-----LEDAIRIAYSDI IYWLVE---DRGPFQKAYWEL 300  
Dy 303 IGPVFRFSENVJFEGISVSGKQHYLDAT-VAYKAVLUNA:DYLFKFGYSKEQVY:LL 362  
Dy 301 SQCG-KVRLGNMVPKVTVAAM 321  
Dy 362 SCCPCGR:SGIVDSPNAVATL 383

RESULT 13  
E71097  
hypothetical protein PH1041 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 18-Aug-2000  
C:Accession: E71097  
R:Kawarabayashi, Y.; Swada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; O  
X.; Onikubo, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; O  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil  
A:Reference number: A71000; MUID:98344137; PMID:9679134  
A:Accession: E71097  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-389 <NA>  
A:Cross-references: GB:AP000004; NID:93236131; PIDN:BA430139.1; PID:dl031082; PID:9  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by Gen  
C:Genetics:  
A:Gene: PH1041  
C:Superfamily: pyrococcus horikoshii hypothetical protein PH1041

Query Match 10.2%; Score 179; DB 2; Length 389;  
Best Local Similarity 22.5%; Pred. No. 3.7e-07;  
Matches 69; Conservative 32; Mismatches 86; Indels 120; Gaps 11;

Dy 88 IMVNGAEKGVLAIVTESN-----LPRGVDPYGCAMLPHEGGLTGTDLTA-----MLND 137  
Dy 16 VAVEGAKVGDATATIRVIRVLSKAASGVDS-----PREGSVGDPYVAKKCPVNE 68  
Dy 138 PIPE-----KVRMTK-----LDSE-----KVWV 155  
Dy 69 PWPEFEVWIGEDAIRCKKCGAPVSPFRMNGYTWVFDHEGGVGVTVYDKKIAEMIAKNW 128  
Dy 156 SKRHITLP-----YKPHIGTISVSPETDSINSLPQNHG----- 188  
Dy 129 -EWHALPKNSKOVPIILYAKADIVCPVSPRIKPIGQLGTIPAVD-----IPDSHNGDFG 182  
Dy 189 -----GNMVDPIGPGSIYPLVRAPGGRLFTGDHACOGDG 225  
Dy 183 YFLINAPHPYAIKEDYETKLTGDHLVDVSVREGAIVIAVPKVEGGVYAGDAHAHEGDD 242  
Dy 226 EICGTAVEFASITTIKVLDIKNNOLSWPRMENAEINMSISGARPLEDATRIAYFDLIYWL 285  
Dy 243 EVAGHTTDTVAETITVEVIGKLNLEGPILLPEEDLP-----PLAKPWKRDEWEKVERL 297  
Dy 286 VEDFGPE 292  
Dy 298 AKKEGIE 304

RESULT 14  
A47696







```
Db 15 EDDQIENRHHPPDLMIAWVKPGDEKVECDMDWIGQIIGNKSDANDVROW--ELIQVHYL 72
QY 82 NPQNGPIMVNGAFKGDVAVYIESMLPRGVDPYVGCAMLIHFQSGJSTDLTAMLDPLP 110
Db 73 ---SGPIGVEAGFCDLWVJLDVGTGFDOSOWS-----FNSLFAKNGSGFLCHFP 122
QY 141 EKVRMIKIDSKVYWSKHHT-----LIPKPHICT-----LAYSPE-175
Db 123 EASKTI-WDFHGVTTSHRVPKRVYAGIMHPCLGCLPSEKLLDINKKREGDILAIUPR 181
QY 176 -----DSTNSJTPDNGHGNMDVDPIDGSGSTIYPLVRA 209
Db 182 VPPIACPTSGSAVMGRSLSDAARAKAAGARTVFPFDHGGNCDDIKNLTGGRVAFVAV 241
QY 209 PGRSLFLGIAHACQDGEI--CGTAVEFASIT--KVLRIKVMOLSWPFEENAKINISIN 246
Db 242 KGGG:SMGDLHFSUGDGEITCG-AIFWAGYLDIKVSEIK-----DGVKKYK 244
QY 267 ARPLEDAITR--AYRDLIYK:VEDFGREO 293
Db 289 KNPVFOFSPITPTVRD--YMIFEGISVDE 315

RESULT 2
AMDA_MYCSM
ID AMDA_MYCSM STANDARD: PRT: 405 AA.
AC Q07838;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetamidase (EC 3.5.1.-).
CS Mycobacterium smegmatis.
CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1772;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-23; 81-100 AND 180-194.
RC STRAIN=NCIC 8159;
RX MEDLINE=93232774; PubMed=8473863;
RA Mahenthalingam E., Draper P., Davis E.O., Colston M.J.;
RT "Cloning and sequencing of the gene which encodes the highly
RT inducible acetamidase of Mycobacterium smegmatis.";
RL J. Gen. Microbiol. 139:575-583(1993).
CC -!- FUNCTION: ALLOWS ACETAMIDE TO BE USED AS A SOLE CARBON OR NITROGEN
CC SOURCE.
CC -!- CATALYTIC ACTIVITY: Acetamide + H(2)O - acetate + NH(3).
CC -!- SIMILARITY: STRONG, TO M.METHYLOTROPHUS FORMAMIDASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X57175; CAA04452.1;
CC DR InterPro: IPR004304; FmdA_AmdA.
CC DR Pfam: PF03069; FmdA_AmdA.1.
CC KW Hydrolase.
FT INIT_MET 0 C POTENTIAL.
SQ SEQUENCE 405 AA; 43933 MW; 05E2E359210F09F2 CRC64;

Query Match
Best Local Similarity 10.2%; Score 178.5; DB 1; Length 405;
Matches 92; Conservative 39; Mismatches 13; Indels 117; Gaps 19;

QY 46 TIEPQDRITVDRFAEGAI-----NSEQDT---PSCILAKPPLPQNGPIMVNGAEGK 95
Db 33 TVKPGSEFIECKENTDQIGNNSANDVROVDIAPCHML-----SGPIRVEGREG 44
QY 97 DVIAYVIESMLP-----RVDVPGICAMIDFEG----- 124
```

```
85 D...IVDILD:GPVQTNGPNCGEWGYSGIFKYNVGGEFLTDDYDPAYKAIWDFHGOQCT 144
125 -----GLTGT-----DLTAMLDNLPPEKVRMIKIDSEKVVWSKRHTLPYKP 165
145 SRHVPGVAYTITHPGLGTAPSPDLAKWNE---RERA--LATDPDRV-----PPLALPP 196
QY 166 HL-GFL--SVSPET-----DSINSLTPDNRHGNMDVDPIDGPGS-ITYPVLRAPGGRKLF 215
Db 197 LDVDTGLGTASGDLLOAIANDGARTVPRENGGHHDIKNFTRGSRFPYP-VFVEGAMLSG 255
QY 216 GDAHACQDGEI--CGTAVEFASIT--IKVDLIKVMOLSWPFEENAKINISIGSARPL--- 270
Db 256 GDLHFSQDGEINFCG-AIEMGGFIDMEVDLIKGMETVGT--NP--IFMPGRVEPLYSE 312
QY 271 -----EGATRIAYRDLIYKVE---DFGFEQWDAYMLL 300
Db 313 WLTFIGISVDHAENNNAYMDAT--MAYRNACLNATLEYLKKWGTGCEQAYLIL 362

RESULT 3
UURA_NEIGG
ID UURA_NEIGG STANDARD: PRT: 950 AA.
AC Q5C968;
DT 31-NOV-1997 (Rel. 35, Created)
DT 31-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exonuclease ABC subunit A.
GN UURA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97340932; PubMed=9197406;
RA Black S.G., Fyfe J.A.M., Davies J.K.;
RT "Cloning, nucleotide sequence and transcriptional analysis of the
RT uvrA gene from Neisseria gonorrhoeae.";
RL Mol. Gen. Genet. 254:479-485(1997).
CC -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UURA IS AN ATPASE
CC AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UURA, UVRB AND UVRC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UURA SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC FMBL: U34760; AAA84855.1;
CC DR InterPro: IPR003439; ABC_transportr.
CC DR InterPro: IPR004602; UvrA.
CC DR Pfam: PF00005; ABC_tran; 2.
CC DR Pfam: PD000006; ABC_transportr; 1.
CC DR TIGRfams: TIGR00630; uvrA; 1.
CC DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
CC KW SOS response; Exonuclease; DNA repair; ATP-binding; Repeat;
CC DNA-binding; Zinc-finger.
FT NP_BIND 42 49
FT NP_BIND 649 656 ATP (POTENTIAL).
FT NP_BIND 262 289 ATP (POTENTIAL).
FT ZN_FING 262 289 C4-TYPE.
FT ZN_FING 748 774 C4-TYPE.
FC SEQUENCE 950 AA; 105552 MW; 359806240AADD6F CRC64;

Query Match
Best Local Similarity 5.7%; Score 99.5; DB 1; Length 950;
Matches 80; Conservative 56; Mismatches 127; Indels 125; Gaps 20;
```



```

RA Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
RA Doclittie W.F., Ragan M.A., Charlebois P.,
RT "Organizational characteristics and information content of an
RT archaeal genome: 156 kb of sequence from Sulfolobus solfataricus
RT P2."
RT Mol. Microbiol. 22:175-191(1996).
RN [3].
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=2133296; PubMed=1142726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awaye M.J., Chan-Weher C.C.Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Pena X.,
RA Charlebois R.H., Doclittie W.F., Duguet M., Gaasterland T.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tostrup N.,
RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Enantioselective, active on 2-to 6-carbon aliphatic
CC amides and on many aromatic amides
CC -!- CATALYTIC ACTIVITY: A monocarboxylic acid amide + H(2)O -> a
CC monocarboxylate + NH(3).
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: Active over the pH range 4-9 and at temperatures
CC from 60 degrees to 95 degrees Celsius.
CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF250611; AAK83092.1; -
DR EMBL: Y08256; CAA59448.1; -
DR EMBL: AE006819; AAK42301.1; -
DR InterPro: IPR000120; Amidase.
DR Pfam: Pf01425; Amidase; 1.
DR PROSITE: PS00571; AMIDASES; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 504 AA; 55855 MW; A9103AB7D09A88EE CRC64;

Query Match 5.6%; Score 97.5; DB 1; Length 504;
Best Local Similarity 21.4%; Pred. No. 1.5;
Matches 54; Conservative 32; Mismatches 29; Indels 77; Gaps 11;

QY 79 PFLPNQCPIMVNGAEKGVAVIESMLPRGVDPYGCICAMIPFGGLIGTDLTLMNDP 138
DB 157 PVLPAPRNEYWAGSSGSAVAV-----ASGYCDMA--VGSDQGS----- 135
QY 139 LPKVRMIKLDSEKV-YMSKRHTLPYKPHIGTSLVSPEIDSN-----SLTPDNHG 188
DB 196 -----TRIPSSWVGIVGLKPTKGLVPTGAFSTETLQGLPMANTVQDVALLEVA 248
QY 189 CNDVDPICGCSITYPLVAPAGGRIFGDAHACGCGEIGCSAVEPASITIKVQLKKN 248
DB 249 GRDELSDQPSLPPPVYK-PYSKLIDG-----VSDMKVGIYKE- 287
QY 249 QLSWPRMENAINISGARPLEDAIRIAYRULIYML-VEQ-----FGFEGQDAFMIL 303
DB 288 GFNNSENKDYDELVDLSAKKLED-----YQVKVEDTSTPLRMGLDITP-AIE 337
QY 301 SOCGKVRIGNMV 312
DB 338 GATATMILGSGV 349

RESULT 6
UVRA_NEIMB STANDARD: PPT: 949 AA.
ID UVRA_NEIMB

```

```

AC 3942P1;
CT 15-JUN-2002 (Rel. 41, Created)
CT 15-JUN-2002 (Rel. 41, Last sequence update)
CT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exonuclease ABC subunit A.
GN UVRA OR NMB0962.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1].
RP SEQUENCE FROM N.A.
RX STRAIN=MC58 / Serogroup B;
MEDLINE=20175755; PubMed=10710307;
RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Gwinn M.L., Deboy R., Pedersen J.D., Dodson R.J.,
RA Nelson W.C., Salzman S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittoni H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiognani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2003).
CC -!- FUNCTION: The ABC excision nuclease is a DNA repair enzyme that
CC catalyzes the excision reaction of UV-damaged nucleotide segments
CC producing oligomers having the modified base(s). Uvra is an ATPase
CC and a DNA-binding protein that preferentially binds single-
CC stranded or UV-irradiated double-stranded DNA (By similarity).
CC -!- SUBUNIT: Consists of three subunits: uvraA, uvraB and uvraC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVRA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF002447; AAF41368.1; -
DR TIGR: NMB0962; -
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR004602; Uvra.
DR Pfam: PF00005; ABC_tran; 2.
DR ProDom: PD000006; ABC_transportr; 1.
DR TIGRFAMs: TIGR00630; uvra; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW DNA-binding; Zinc-finger; Complete proteome.
FT NP_BIND 42 49 AIP (POTENTIAL).
FT NP_BIND 649 656 AIP (POTENTIAL).
FT ZN_FING 262 285 C4-TYPE.
FT ZN_FING 748 774 C4-TYPE.
SQ SEQUENCE 949 AA; 105536 MW; 153F5AE2D48CDD35 CRC64;

Query Match 5.6%; Score 97.5; DB 1; Length 949;
Best Local Similarity 20.4%; Pred. No. 3.5;
Matches 75; Conservative 57; Mismatches 127; Indels 125; Gaps 20;

QY 12 RGVAGRKPVTHTLTEMKKEFHTTIGYTPVLTIEPGDRIIVDTRDAPEGAINSDI 71
DB 544 RDLGNSVIVVEHDEDAIREADFVDMGPGAG-----EHGGNVL--ADTPENVAQCNSV 596
QY 72 PSQLL-----KMPFLNPONGPIVNGAEKGVAVIESMLPRGVDPYGCICAM 120
DB 547 TGOYLSGKKSIAPVSEKTP-VNPDRLVVLKAGRNKLVNLTLE--LPLG-----LI 644
QY 121 PRFGGLTGLTAMNDPLPEKVRMIKLDSEKVYMSKRHTLPYKPHIGTSLVSPEIDSN 180
DB 645 TCITGVSGSGKSTLINDTL-----AKITARELNRAQEAPF-----DDIH 685

```

```
QY 181 SLTDNNGGMDVPCIG-----PGSIT-----YFLVRAFG---SKLFI --- 215
D 686 GLEHLDKVINVDOSPGRTPRSNPNATYGLFTTIRELFGVPLSRGNGVNGWSPFNKVG 745
QY 216 GDAHACOGDG-----EIC-----GTAVE--- 233
D 745 GREACOGGCV:KVENHFLPDVYVPOCEVCHGKRYNRETLETQYKGNISQVLMYFEAR 605
QY 234 --FASITTI--KYDLINKWOLSWPRMENAENIMSIGSARPLEDATRIAYRD---LIIYMW 246
D 806 EFFDAVTVSRKLGJLMDVCGGIVRLGOSATTLSGCGAQRVKLAELSRRGTGRTLIYLD 965
QY 287 E-DGFFQWDAYMLISCGKVR-LGNMV 312
D 866 EPTTGLHFADIALLELVIGRLKKGKNSI 893

RESULT 7
GALT_THETN
ID GALT_THETN STANDARD; PRT: 519 AA.
AC QBR86;
DT 15-JUN-2002 (Rel. 41, Created);
DT 15-JUN-2002 (Rel. 41, Last sequence update);
DE Galactose-1-phosphate uridylyltransferase (EC 2.7.7.10) (UniProt)
DE uridylyltransferase.
GN GALT OR TFE1929.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21592816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:889-730(2002);
CC -!- CATALYTIC ACTIVITY: UDP + alpha-D-galactose 1-phosphate =
CC -!- diphosphate + UDP-galactose.
CC -!- PATHWAY: Galactose metabolism; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasm; (potential).
CC -!- SIMILARITY: BELONGS TO THE GALACTOSE-1-PHOSPHATE
CC -!- URIDYLTRANSFERASE FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: AEO13143; ANM25106.1;
CC FR PROSITE: PS01163; GAL_P_UDP_TRANSF_II; FALSE_NEG.
CC KW Transferase; Nucleotidyltransferase; Galactose metabolism;
CC KW Complete proteome.
CC SEQUENCE 519 AA; 50467 MW; 0BF788A4CA08ED75 CRC64;

Query Match 5.43; score 94.5; DB 1; length 519;
Best Local Similarity 20.28; Pred. No. 2.8;
Matches 56; Conservative 42; Mismatches 94; Indels 85; Gaps 13;

QY 83 PONGPIMVNGAEXGDVLAVIESMLPRGVDPYGLICAMIPHFGSLGTDLIAMLNWLPK 142
D 158 PEKDPKEIEKAKK-----IPO--SGYKCLGCIENVGFGN-----LNHPARON 139
QY 143 VRMIKLD-----SEKVVSKRHTLPYKPHUIGTSLVSPE-----IDSINSLTPDNH 187
D 200 LRITPVKVGQWYQFYSPVYVYNEHCILLHEPHI-PMKISEKTFVRLDFIDDFHYFM 258
```

```
QY 188 GGNMDVDPDUGGSI-----TYPLVKAHGGRIFGDAHACXGDSGICGIAVEFAS 236
D 259 GSNADLPIVG--GSLVHFHFQGGHTEPMEEAPEEYFHPKY----- 360
QY 237 IITTKVLLINKWOLSWPRMENAENIMSIGSARPLEDATRIAYRDLIYWLVDGFGFWDA 296
D 301 -KEYKAGTLK-WPMSVIRI-----SSKOREKLTLSHILNWL--KGYSDESVDV 346
QY 297 YMLLSQCGKVRLGNMVDP-----KYTVGAMLNKN 325
D 347 LAYSEKDGKIVPHNTITPIARFNKEGEYEIDLVLRNN 363

RESULT 8
NIA_ASPNG
ID NIA_ASPNG STANDARD; PRT: 867 AA.
AC P36858;
DT 01-JUN-1994 (Rel. 25, Created);
DT 01-JUN-1994 (Rel. 29, Last sequence update);
DT 01-NOV-1995 (Rel. 32, Last annotation update);
DE Nitrate reductase [NADPH] (EC 1.6.6.3) (NR).
GN NIAH.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92175518; PubMed=1541396;
RA Jaxles S.E., Campbell E.L., Punt P.J., Hawker K.L., Contreras R.,
RA Hawkins A.R., van den Hondel C.A.M.J.J., Kinghorn J.R.;
RT "The Aspergillus niger niaD gene encoding nitrate reductase: upstream
RT nucleotide and amino acid sequence comparisons.";
RL Gene 111:149-155(1992);
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -!- CATALYTIC ACTIVITY: NADPH + nitrate = NADP(+) + nitrite + H(2)O.
CC -!- COFACTOR: REQUIRES FAD, A HEME GROUP (CALLED CYTOCHROME B-557)
CC -!- AND ONE MOLYBDENUM ATOM.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC -!- N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC -!- C-TERMINAL DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: M77022; -; NOT_ANNOTATED_CDS.
CC FR PIR: JQ1525; JQ1525.
CC FR HSP: P17571; 2CND.
CC DR InterPro: IPR001199; Cyt_B5.
CC DR InterPro: IPR001834; Cyt_B5_reductase.
CC DR InterPro: IPR000572; Euk_Mb_oxred.
CC DR InterPro: IPR005066; Mo-co_dimer.
CC DR InterPro: IPR001433; Oxred_FAD/NAD(P).
CC DR Pfam: PF00173; heme_1; 1.
CC DR Pfam: PF00174; oxidored_molyb; 1.
CC DR Pfam: PF00970; FAD_binding_6; 1.
CC DR Pfam: PF03404; Mo-co_dimer; 1.
CC DR PRINTS: PR00406; CYTB5REDTASE.
CC DR PRINTS: PR00363; CYTOCHROME_B5.
CC DR PRINTS: PR00407; EUMOPTERIN.
CC DR PRODOM: PD000612; Cyt_B5; 1.
CC DR PROSITE: PS00191; CYTOCHROME_B5_1; 1.
CC DR PROSITE: PS50255; CYTOCHROME_B5_2; 1.
```

```

DR PROSITE: PS00559; MOLYBDOPTERIN_EUK_1;
KW Oxidoreductase; Flavoprotein; FAD; NADP; Heme; Molybdenum;
FT METAL 152 152
FT METAL 203 203
FT DISULFID 395 399
FT DOMAIN 518 598
FT BINDING 549 549
FT BINDING 572 572
FT DOMAIN 612 567
FT NP_BIND 937 846
SQ SEQUENCE 867 AA: 97188 MW: 8540335DCBE12E42 CRC64;

Query Match
Best Local Similarity 20.7%; Pred. No. 6;
Matches 63; Conservative 29; Mismatches 95; Indels 118; Gaps 13;

QY 88 IMVNGAEKGDVAVYIESMIPRGVDPYVIGC-----AMIPHEG-----GLTSDITAKEND 137
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 VMEGAOK-----LFGVY--YGTSLKLNKADPNPGIMLAHKMKGSEMERFURK 248
QY 138 PLPKKVKIKLDEKVVYWSKHTLP-----YKPHCTLSVSPEDSI----- 179
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 249 PL-RAVVPQIGRSVAKWLKLLIDAPSNNWYHYDNRVLPTNVSPESSSDPNKWKOD 307
QY 180 -----NSLPQWNGSNMVDPIGPGSITYPLVRAPEGRFLFGDAHACCGGFEIC 323
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 308 RYATVDLNVSSVYVPEKEVLDLASAGPSVNVGYAYAGGRPT----- 352
QY 230 TAVEFASITIKVDLKNWQL-----SW----- 352
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 353 TRVE-----ISLDKGSWRLANISYAEKRYDFEGDLFGGRVHMSWRETCFCWCEWSLD 406
QY 253 ---PRMENAEMISGSRPLEDATRAYRLIYVLVEDFGFEQWDAYMLLSOGCKVRLG 309
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 407 IAIPELENTDAII-----VRANDEALALQPRDM-YWSVLGMNPNPFRVTIKENTLRFE 481
QY 310 NWVDP 314
DB : |||
DB 462 HPTDP 466

RESULT 9
CPUL_RABIT
ID CPUL_RABIT STANDARD: PRI: 500 AA.
AC P52786;
DT 01-OCT-1996 (Rel. 34, Created);
DT 01-OCT-1996 (Rel. 34, Last sequence update);
DT 15-DEC-1998 (Rel. 37, Last annotation update);
DE Cytochrome P450 2J1 (EC 1.14.14.2) (CYP1J1) (P-4501B).
GN CYP2J1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Cricetidae;
CX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92011499; PubMed=1717443;
RA Kikuta Y., Sogawa K., Hanai M., Kinoshita M., Kusunose E., Nojima Y.,
RA Yamamoto S., Ichihara K., Kusunose M., Fujii-Kuriyama Y.;
RT "A novel species of cytochrome P-450 (P-4501b) specific for the small
RT intestine of rabbits: cDNA cloning and its expression in COS cells.";
RJ J. Biol. Chem. 266:17821-17825(1991).
CC -:- FUNCTION: CATALYZES THE N-DEMETHYLATION OF BENZPHETAMINE TO
CC FORMALDEHYDE.
CC -:- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) -> ROH +
CC oxidized flavoprotein + H(2)O.
CC -:- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -:- TISSUE SPECIFICITY: SMALL INTESTINE.
CC -:- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation .

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMH: D90405; BAAL4401.1;
DR HSP: P00179; LD76.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450.1.
DR PRINTS: P00385; P450.
DR PROSITE: PS00666; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT INIT_MET 0
FT BINDING 446 446 HEME (BY SIMILARITY).
SQ SEQUENCE 500 AA: 57194 MW: A0EE628E5FF23DE CRC64;

Query Match
Best Local Similarity 22.3%; Pred. No. 4.3;
Matches 81; Conservative 54; Mismatches 130; Indels 98; Gaps 21;

QY 3 WLEB---SIMARGVAGKPV-----THLTHEMOKE-----PHYTI-GPYSTPLV 45
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 WKQRRPALMTLRNFGLGKKSLEERIQEEARHLTEAMEKEGGQPFDAHFKINNAVSNIC 192
QY 46 IIEPDRI-----IVDRDAFEGAINSEQDIPSOALK-----MPFL-----NPCN 85
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 STTFSEFYHUGQOQLLKLFDEWYLEASMLQLYNIFWIMKFLPCAHQOTLFSNWK 252
QY 86 GPIMV-----NGAEKGDV-JAVYIESM--LPRGV-----DPYICAMIPHGGLT 127
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 253 LELFVSRMLENHKKDWNPAETRDFTDAYLKEMSKYKPGSATSFSENEELICSTLDLF--LA 310
QY 128 GYDLTA-----MENDPLPEKVRMVKLDSEKVV-----SKRHTLPYKPHIGILSV 173
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 GTTSDMRWGLLFMALYPEIOEKVH-AEIDSVICWQCPMSASRESLPY-----TNVAVI 364
QY 174 PEIDSINSLTPDHGNGMDVPDIPGHSITYPLVRAFGGRFLFGDAHACCGGGEICGTAVE 233
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 HEVORMGNILP-----LNVPREVTVDITLAGYHLPKGVVLTNLTLAKKDPE-----E 412
QY 234 FASITTIKVD-LIKWQLSWPRMENAEMISGSRPLEDATRAYRLIYVLVEDFGFEQWDAYMLLSOGCKVRLG 289
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 413 WATPDTPNPEHFLNGOF---KKKFAFTPFSSICKRACLG--QLAKSELFIFFTSLMKRF 467
QY 290 GFE 292
DB : |||
DB 468 TFK 470

RESULT 10
ABFA_HACSU
ID ABFA_HACSU STANDARD: PRI: 500 AA.
AC P94531; O05096;
DT 01-NOV-1997 (Rel. 35, Created);
DT 01-NOV-1997 (Rel. 35, Last sequence update);
DT 15-JUN-2002 (Rel. 41, Last annotation update);
DE Alpha-L-arabinofuranosidase 1 (EC 3.2.1.55) (Arabinosidase).
GN ABFA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
CX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=97217725; PubMed=9084180;
RA Sa-Noqueira I.M.G., Noqueira T.V., Soares S., de Lencastre H.;
RT "The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence,
RT genetic organization and expression.";
RJ Microbiology 143:957-969(1997).
RN [2];
RP SEQUENCE FROM N.A.

```

```

HQ      SEQUENCE      506 AA:  57060 MW:  7397ED52A4987686 CRC64:

Query Match:          5.2%:  Score 90.5;  DB 1;  Length 500;
Best Local Similarity 25.9%:  Pred. No. 5.7;
Matches 52;  Conservative 32;  Mismatches 74;  Indels 43;  Gaps 13;

QY      14:  EKVRMIKELDSE-----KVVWSKRRHTLPYKPHGTLSVSLFSDINSITPDNHGNGMD 152
DB      :  ||| ||| :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      2  KAKRMI-VDKYEKIGEDVKRIYS-----FIIHNGR-AVTEGIVPEOHPEADEGFRKD 53
DB      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      193 VPDIGCSITVPLVRAPGGRIFGDARACQGGDEICG--TAVEPASITITKVDLIKNNQL 250
DB      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      54  VQSL-IKELQPIRIVDGGN-FISGYWWEKVGVPENRPRDLIAWQTETNEVGTFNEFL 111
DB      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      251 SWPRMKNENIMSIG-SARPLEDATRIAYRDLI-----YW--IVEDFGFQSDWAYM 298
DB      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      112 SWPKKVNTEVNMAYNLCTRGID-----AARNLVEYCNHPKGSYNSDLRSHSGYEQ--PYG 164
DB      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      299 LLSGGKVRKGNMYDPKYTVG 315
DB      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      165 IKTWC----LGNEMDGPWQIG 181
DB      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 11
FTH1_HAETN
ID      FTH1_HAETN      STANDARD;      PRT;      635 AA.
AC      271377;
DT      01-NOV-1997 (rel. 35, Created)
DI      01-NOV-1997 (rel. 35, Last sequence update)
DE      15-JUN-2002 (rel. 41, Last annotation update)
DE      Cell division protein ftsh homolog 1 (EC 3.4.24.-).
GN      FTSH-A OR H11335.
OS      Haemophilus influenzae.
CC      Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
CC      Haemophilus.
CC      NCBI_TaxID=727;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=RD / KW20 / ATCC 51907;
RC      MEDLINE=95150630; PubMed=7542800;
RA      Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA      Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA      McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA      Scott C.D., Shirley R., Liu L.T., Glodek A., Kelley J.M.,
RA      Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA      Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA      Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA      Gachm C.L., McDonald B.A., Small K.V., Fraser C.M., Smith H.O.,
RA      Venter J.C.;
RA      "Whole-genome random sequencing and assembly of Haemophilus influenzae
NT      Rd.";
RI      Science 269:496-512(1995).
RI      [2]
RE      REVISIONS.
EL      White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.;
EL      Submitted (SEP-1996) to the EMBL/GenBank/DBRI databases.
CC      !- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE
CC      (BY SIMILARITY).
CC      !- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC      !- SUBCELLULAR LOCATION: Integral membrane protein.
CC      !- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC      !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      EMBL; U32812; AAC22979.1;
CC      TIGR; H11335;

```



```
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003599; AAA_ATPase_centr.
DR InterPro: IPR003960; AAA_sub.
DR InterPro: IPR000642; Peptidase_M41.
DR Pfam: PF01434; AAA; 1.
DR SMART: SMC0382; Peptidase_M41; 1.
DR TRICRAMS: TIGR01241; PtsH_fa; 1.
DR PROSITE: PS00674; AAA; 1.
KW Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;
KW Zinc; Complete proteome.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 24 POTENTIAL.
FT DOMAIN 25 95 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 96 120 POTENTIAL.
FT DOMAIN 121 635 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 191 198 ATP (POTENTIAL).
FT METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 414 417 BY SIMILARITY.
FT METAL 417 417 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 635 AA: 70030 MW: 70E6DD1E293839D CRC64;
Query Match 5.18; Score 89.5; DB 1; Length 635;
Best Local Similarity 20.48; Proc. No. 9.4;
Matches 80; Conservative 59; Mismatches 117; Indels 143; Gaps 20;
QY 13 GVAGAKPVTHHLEEMKEFHNTICPYSTFVLTIEGDMHIVLTQAF--- --53 63
Db 226 GVASR---VRDMEQAKK-----NAPCI-----IFIDEIVAVKRGAGLGG 56
CY 64 AINSEDIPLSKMPPFNQNGPIKVNKAKEGV----- 98
Db 267 HDREQTLQCLMVEGDFSGNCGVIVIAATNREFVLPALTRGFEURQVYVGLFVVKGR 326
QY 59 ---LAVVIESN-LPFGVDPYGCAMPHFGCLGIDLTAMLNPL----- 139
Db 327 EQILKVMKRVSAQVDVAMTLARGTP---GYSGADLANLVAAALFAAVNRKRTVINE 183
QY 140 -----PEKVRMIKLDSEK---VYMSKRHILPKPHICITLSVSPEDIS:NSLIP 185
Db 384 FEKAKCKINMGPERITMTOKESTAYHEAGHAI-----VGVL--VPEDVPVHVT-- 434
QY 186 NHGGMNDVDPGISTYPLVRAPGGRFLFGDAHACGGGEICGTAVEFASITTIKVDLI 245
Db 435 -----IPRGRALGVTFPLP-----EGD-QTISQKQLES-----KISTL 468
QY 246 KMWLSHPRMENAIMSISARPLEDATRIAYRDLIYWLVEDFGP-EQWDAYMLLSQG 304
Db 469 YAGRLAEDLIYGENI-STGASNDIKVATNIARNMVTQW-----GFSEKGLPILYTEDEG 522
QY 305 KVRIG-NMVPKY-----TVGAMLNKN 325
Db 523 EVFLGRSKAKAKHMSDETAHSIDEVRAIVNRN 553
RESULI 12
ID PKK_ECOLI
AC P28688; Q47549; STANDARD; PRT; 687 AA.
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Polyphosphate kinase (EC 2.7.4.1) (Polyphosphoric acid kinase) (ATP-
DE phosphate phosphotransferase).
GN PPK OR B2501 OR Z3764 OR ECS3363.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-K12;
KX MEDLINE=93054553; PubMed=1331061;
RA Akiyama M., Crooke E., Kornberg A.;
RT "The polyphosphate kinase gene of Escherichia coli. Isolation and
RT sequence of the ppk gene and membrane location of the protein.";
RL J. Biol. Chem. 267:22556-22561(1992).
KN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.E., Rode C.K., Mayhew G.F.,
RA Gregor C., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
KN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97345980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isoro K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizutuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel J., Satch Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takenoto K., Uehara K., Wada C.,
RA Yagagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:191-111(1997).
KN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL53 / AFCC 700927;
RX MEDLINE=21074935; PubMed=1120652;
RA Poria N., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Gerbeek E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
KN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida I., Takami H., Honda T., Sasakawa C., Shinagawa H.,
RA Kahara S., Shiba T., Hattori M., Ogasawara N., Yasunaga T.,
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
KN [6]
RP SEQUENCE OF 1-90 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=87260114; PubMed=3301838;
RA Smith J.M., Daum H.A. III;
RT "Identification and nucleotide sequence of a gene encoding 5'-
RT phosphoribosylglycinamide transformylase in Escherichia coli K12.";
RL J. Biol. Chem. 262:10565-10569(1987).
KN [7]
RP SEQUENCE OF 678-687 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=93107072; PubMed=8380170;
RA Akiyama M., Crooke E., Kornberg A.;
RT "An exopolyphosphatase of Escherichia coli. The enzyme and its ppk
RT gene in a polyphosphate operon.";
RL J. Biol. Chem. 268:633-639(1993).
KN [8]
RP SEQUENCE OF 433-435, AND MUTAGENESIS OF HISTIDINE RESIDUES.
RX MEDLINE=9721395; PubMed=8962061;
RA Kumbic K.D., Ann K., Kornberg A.;
RT "Phosphohistidyl active sites in polyphosphate kinase of Escherichia
```

```

RT  coli."
RI  Proc. Natl. Acad. Sci. U.S.A. 93:14391-14395(1996).
CC  -!- FUNCTION: Catalyzes the reversible transfer of the terminal
CC  phosphate of ATP to form a long-chain polyphosphate (polyP).
CC  Can form linear polymers of orthophosphate with chain lengths up
CC  to 1000 or more. Can also act in the reverse direction to form
CC  ATP in the presence of excess ADP. Can also use GTP instead of
CC  ATP, but the efficiency of GTP is 5% that of ATP.
CC  -!- CATALYTIC ACTIVITY: ATP + (phosphate)(N) = ADP + (phosphate)(N+1).
CC  -!- SUBUNIT: HOMOTETRAMER.
CC  -!- SUBCELLULAR LOCATION: Inner membrane-associated
CC  -!- P.M. An intermediate of this reaction is the autophosphorylated
CC  ppk in which a phosphate is covalently linked to histidine
CC  residues through a N-P bond.
CC  -!- SIMILARITY: BELONGS TO THE POLYPHOSPHATE KINASE FAMILY.
CC  -!- SIMILARITY: CONTAINS 1 PLD PHOSPHOTRANSFERASE DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: L03719; ; NOT_ANNOTATED_CDS.
DR  EMBL: AC00336; AAC7554.1; .
DR  EMBL: D90878; BAA16389.1; .
DR  EMBL: AF005479; AAG57611.1; .
DR  EMBL: APC02551; BAB36786.1; .
DR  EMBL: M13747; AAA03903.1; ALT_INIT.
DR  EMBL: L06129; ; NOT_ANNOTATED_CDS.
DR  F01: B28486; B28486.
DR  F01: A44306; A44306.
DR  EcoGene: EG1510; ppk.
DR  InterPro: IPR001736; PJD.
DR  InterPro: IPR003414; PP_kinase.
DR  Pfam: PF02503; PP_kinase; 1.
DR  PROSITE: PS50035; PLD; 1.
KW  Transferase; Phosphorylation; Inner membrane; Complete proteome.
FT  INIT_MET 0 6
FT  DOMAIN 429 403 PLD PHOSPHOTRANSFERASE.
FT  ACT_SITE 434 434 FORMS THE PHOSPHOTRANSFERASE INTERMEDIATE.
FT  ACT_SITE 453 453 FORMS THE PHOSPHOTRANSFERASE INTERMEDIATE.
SQ  SEQUENCE 687 AA; 80300 MW; D040A2AD3D725804 CRC64;

Query Match 5.18; Score 89.5; DB 1; Length: 667;
Best Local Similarity 23.5%; Pred. No. 10;
Matches 47; Conservative 39; Mismatches 69; Indels 45; Gaps 12;

QY 25 LTRKQKEPHYTYGYPV---LTIFPGD--RIIVDTIRDAFCAINSEQIIPSGE--LK 77
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 479 ITNEVRVFNFIENPY-RVTFEDYLMVNSPONSRELLIEMVD--RIANAQGLPSGILK 535

QY 78 MFLNPNQPIWVNGAEKGVAVIESM--LPRGVDPVGIQAMIPHGGLTGDTLAML 135
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 536 ---LN-----NLVKGGLVRLAASSGSPVNLVSGMCS;PNEGISDNIRAI 564

QY 136 NDLPEKVRMKIDS---EKVY-----NSKRFTLPVXPHGLTSLVSPEI----- 176
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 585 VDRYLEHDRVYFFNGGDKKKVSSADMMTRK-IDYRIEATVPLDPLKORVILITDIL 643

QY 177 --DSINSLTPDNHGGNMVDP 194
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 644 FSDTVARVYDKELSNRYVF 653

RESULT 13
ID IGA4_HAEIN STANDARD; PRT: 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```

```

DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA.
OS Haemophilus influenzae.
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NHII HK61;
RX MEDLINE=92234949; PubMed=1373717;
RT "A comparative genetic study of serologically distinct Haemophilus
R: influenzae type 1 immunoglobulin A1 proteases.";
RC J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: M87491; AAA24968.1; .
DR  MEMOPS: S05.001; .
DR  InterPro: IPR000710; Iga_S6.
DR  InterPro: IPR004899; Pertactin_sup.
DR  Pfam: PF02395; IGA1; 1.
DR  Pfam: PF03212; Pertactin; 2.
DR  PRINTS: PR00921; IGASRPRTASE.
KW  Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT  SIGNAL 1 25 POTENTIAL.
FT  CRAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT  PROSEP :022 1849 HELPER PEPTIDE (POTENTIAL).
FT  ACT_SITE 299 299 PROBABLE.
SQ  SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 5.0%; Score 88; DB 1; Length 1849;
Best Local Similarity 19.0%; Pred. No. 49;
Matches 73; Conservative 47; Mismatches 128; Indels 136; Gaps 20;

QY 35 YTGPISTPVLITPGD---RIIVDTIRDAFE-GAINSE-ODIPSOLLAMPFLNPONGPI 88
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17 YALPTPTAALVDRDOVDYQIFRDAENKGFSGATNVEVRKKNSL-----GSA 67

QY 89 MVNCAEKGDVLAIVIESMLPRGVDPV-----GICAMIPHFGCLGTDLTAMLNDP 138
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 LPRGIPMIDFSDVDVKRIATLVNPQVVGKVKVNGVSEL--HFNGLNGN-----MNG 120

QY 139 LPEKVRMKILISEKVYNSKRHLTP-----YKPHIGTSLVSPEI 176
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 121 NAKSHRDVSEENRYITVEKNFTEPNTVSTTTKEQDAQRREDYVMPRLDKFTV--EV 178

QY 177 DSINSLTPDNHGGNMVDPIDGPSIYTP-LVRAPGGRLFIGNAHACQGGGICGTAVETA 235
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 APTEASTANNKGEYNNSD-----KYPAPVRLGSGSQFIYKGS-----RYQ 220

QY 236 SITTIK---VDLIKNNLSWPRME-----NAENIMTSGSARPLE---D 272
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 221 LILTEKDKQGNLJRNWDVGGDNLVGNATYTGIACTPYKVNHNGLIGFGNSKEHSD 280

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 15:44:52 : Search time 80 Seconds  
(without alignments)  
844.793 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1748  
Sequence: 1 MKWDEESIMAKRGVGAGRKPR.....GNMVDPRYTVGAMLNKLNLY 328

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL21:  
1: sp.archaea:  
2: sp.bacteria:  
3: sp.fungi:  
4: sp.human:  
5: sp.invertebrate:  
6: sp.mammal:  
7: sp.mhc:  
8: sp.ordanelle:  
9: sp.phage:  
10: sp.plant:  
11: sp.podent:  
12: sp.virus:  
13: sp.vertebrate:  
14: sp.unclassified:  
15: sp.virus:  
16: sp.bacteriap:  
17: sp.archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	368	21.1	317	16	Q92P43
2	363	20.8	300	16	Q9KGN3
3	355.5	20.3	299	16	Q8R8S5
4	348	19.9	344	17	Q8U1V6
5	342	19.6	318	17	Q9V26
6	332.5	19.0	377	17	Q9VEQ1
7	327	18.7	318	17	Q976P0
8	322.5	18.4	304	16	Q9A553
9	306.5	17.4	298	17	Q9V075
10	300.5	17.2	285	16	Q9WXX3
11	269	15.4	328	16	Q8Y1Y6
12	261.5	15.0	299	16	Q9Z7N6
13	239.5	13.7	411	3	Q9C453
14	233.5	13.4	410	3	Q9URV7
15	233	13.3	296	16	Q8V452
16	208	11.9	432	10	Q9SZE9

17	198	11.3	454	10	Q34DX1
18	133.5	11.1	435	16	Q8RCU9
19	166	10.6	452	10	Q94CA8
20	173	10.2	489	17	Q587S5
21	167.5	9.6	382	10	Q94AP4
22	130	7.4	307	10	Q9SZF0
23	107	6.1	238	16	Q8Y6A5
24	105	6.0	842	10	Q9FLJ8
25	104	5.9	344	3	Q96UW2
26	102	5.8	238	16	Q92AL8
27	100.5	5.7	471	3	Q9R634
28	99.5	5.7	501	17	Q8U4J7
29	99	5.6	977	16	Q9ENV6
30	98.5	5.6	348	16	Q9JUS4
31	98	5.6	996	4	Q8TD25
32	97.5	5.6	949	16	Q9JZP1
33	97	5.5	380	3	Q9V772
34	96.5	5.5	432	16	Q97MG8
35	96	5.5	945	17	Q8ZXT6
36	94.5	5.4	470	16	Q8RZM1
37	94.5	5.4	519	16	Q8R8R6
38	94.5	5.4	613	16	Q9ZL36
39	94	5.4	344	3	Q96VR2
40	94	5.4	655	16	Q9RYM2
41	93.5	5.3	466	16	Q8XNM0
42	93	5.3	682	3	Q9R8S3
43	92.5	5.3	505	17	Q9V2G8
44	92.5	5.3	513	17	Q8TUQ7
45	92	5.3	547	2	Q9E1I1

#### ALIGNMENTS

RESULT :

Q92P43 PRELIMINARY: PRT: 317 AA.  
 10 Q92P43  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45

Q94dx1 oryza saliv  
 Q8rcu9 thermoanaer  
 Q94ca8 arabidopsis  
 Q587s5 pyrococcus  
 Q94ap4 arabidopsis  
 Q9szf0 arabidopsis  
 Q8y6a5 listeria m  
 Q9flj8 arabidopsis  
 Q96uw2 trichoderma  
 Q92al8 listeria in  
 Q9p634 neurospora  
 Q8u4j7 pyrococcus  
 Q9env6 thizobium  
 Q9jus4 neisseria  
 Q8td25 homo sapien  
 Q9jzpi neisseria m  
 Q9v772 emicella  
 Q97mg8 clostridium  
 Q8zxt6 pyrobaculum  
 Q8y2ml anabaena sp  
 Q8r8r6 thermoanaer  
 Q92l36 thizobium m  
 Q96vr2 trichoderma  
 Q9rym2 deinococcus  
 Q8xnm0 clostridium  
 Q9p8s3 aspergillus  
 Q9v2g8 pyrococcus  
 Q8tuq7 methanosarc  
 Q9etil citrobacter

Query Match 21.1%; Score 368; DB 16; Length 317;  
 Best Local Similarity 32.4%; Pred. No. 9.2e-24;  
 Matches 102; Conservative 43; Mismatches 128; Indels 42; Gaps 10;  
 Qy 23 HLLTEENKHEFYHTIGYTPVLTIEPFGDRIIVDTDAFEAGINSEQDIPS-QLIKMPL 81  
 DB 26 HH-----HYGWBHSIPPVETVAPGSRLEFNCLDSGSGHGTADSTVADYVDFTKV 66





DT	01-JUN-2001 (TRENBLrel. 17, Created)
DT	01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE	Amidase-related protein.
GN	CC26i1.
OS	Caulobacter crescentus.
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC	Caulobacter.
OX	NCHI_taxid=155892;
RP	SEQUENCE FROM N.A. CB15;
RC	STRAIN=ATCC 19089 /
RC	MEDIAN=21173698; PubMed=11259647;
RA	Nierman W.C., Feldblum I.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA	Potocka J., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA	Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA	Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT	"Complete genome sequence of <i>Caulobacter crescentus</i> ."
RT	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

```

OC.
CAHIOABCTER.  OX
NCBI_taxID=155892:
[1]
SEQUENCE: FROM N.A.
STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
CA Nierman W.C., Feldblyum I.V., Laub M.T., Paulsen I.T., Nelson K.E.,
KA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka K., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBooy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ulsterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
BL PNAS Acad Sci U.S.A. 98:4136-4141(2001).
EMBL: AC005925; AAK24580.1;
DR TIGR: CC2611;
DR InterPro: IPR0043C4; FmdA_AmdA.
DR Pfam: PF03059; FmdA_AmdA; 1.
KW Complete proteome.
SQ SEQUENCE 304 AA; 32646 MW; E31FDAB7E69FD4BF CRC64;

```

```

D0 6S JAGAFBFGDILSVRIEATFPN--NDWGYCAVKP-----LAGT-----LPDEFKRYVYSHAV 111
QY 149 DSE---KVVYASKRHILPYKPHIGTITLSP--EIDSINSLTPDHHGNNMDVPDIPGSGIT 202
D0 114 DQAGOVCKPEKGP--JPLAPFFGTMGVAPKAYGRLLSSREPHEGNNMDNKELVAGSTI. 171
QY 203 YPLVRAPGGRFLFIGDAHACOGDGRICGTAVEPASITTIKVDLIKNQ-----LSWPRM 255
D0 172 YLPVWPVPGANTSVGNGHGRQGDGVCVNALEMGLTGCTFTFVLHKKANGAADTAPFAWPR 231
QY 256 ENAENINSIGSRPLEDATRIAYRDLIYWLVEDFGEOMDAYMLLSQCKVKVLRGNMVDPK 315
D0 232 ETPTHVLMGFHEDLDLAKMQALPQTIDF.TARSNLTRVQAVQFCSLAVDFRVTQTVINCE 291
QY 316 YTVGAMLNKNLL 327
D0 292 KGVHALLSKGLL 303

```

CC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
CC Pyrococcus.  
CX NCBI\_taxID=29292;  
RN [1]  
RP SEQUENCE FROM N A



```

RC STRAIN-ORSAY;
RA Helicobacter;
RT "Prokaryotic genome sequence: Insights into archaeal chromosomes:
KL structure and evolution.";
DR Submitted (2001-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248285; CAP49430.1;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA.
DR Hypothetical protein: Complete proteome.
SQ SEQUENCE 298 AA: 32456 MW: 475843336222 CRC64;

Query Match 17.4%; Score 304.5; DB 17; Length 299;
Best Local Similarity 30.7%; Pred. No. 2.8e-18;
Matches 94; Conservative 45; Mismatches 120; Indels 47; Gaps 10;

QY 35 YTGIPSTPVLTIPEGDRIVDTROAFEGAINSEQDIPSQLKMPF--LNPONGIMVNG 92
DB 14 YSGPNKKEVARAKGEIVFOTLDALGQVKSFE--T:EKIDFSRVNPTGPGVYEG 70
QY 93 AEKGOVLAVYIESMLPRGVDYXGICAMIPHFGLTGTDITAMINDPLPEKV-----RMIX 147
DB 71 AKRGGLRLVDLLDKVEG-----KGVVTPAGAGVGGKVKVERPOTRICE 114
QY 148 EDEKVVWYKRRHLPYKPHIGTILSVSPEDSINSLTPDNHGNMVDYDIPGSGITYPLVR 207
DB 115 VKGCVIF-KGKIPAMPKMGVIGVAYD-EEVPTGPGKHGGMNDINLRKGTTFYFVF 172
QY 208 APGRFLIGDAHACOGDGEICGTAVEFAFASITIKVDLIKNNQSWPRMENAINSIGRA 267
DB 173 VDGAYLAIGDLHVMGGEVGSACEVSGEV-VRVTPMEG-KLENPLLTIEDSEFLVSD 231
QY 268 RPLDTRATRYARDLIYWLVEDEGFE-----QWD-AYMLISQCGKVRGLNMVDPKIV 318
DB 232 ENLKAIEAV-----SLGVEALRKSNDISMDFAIMYASLVNDVEISLVDPKIV 282
QY 319 GAMLNK 324
DB 263 RVRIPK 288

RESULT 10
Q9WXX3 PRELIMINARY; PRI: 285 AA.
AC Q9WXX3:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
LT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Acetamidase, putative.
GN TM0119.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotocaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Neider W.C., Ketchum K.A.,
RA McDonald L., Otterback T.R., Malek J.A., Linher K.D., Garrett M.K.,
RA Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson C.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White G.,
RA Salzberg S.L., Smith H.G., Venter J.C., Fraser C.M.
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999)
DR EMBL: AE001698; AAC35213.1;
DR TIGR: TM0119;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA.
DR Complete proteome.
SQ SEQUENCE 285 AA: 31196 MW: 131E9167BD6DDDBE CRC64;

Query Match 17.2%; Score 300.5; DB 16; Length 285;

```

```

Best Local Similarity 29.7%; Pred. No. 5.8e-18;
Matches 85; Conservative 53; Mismatches 125; Indels 33; Gaps 9;

QY 35 YTGIPSTPVLTIPEGDRIVDTROAFEGAINSEQDIPSQLKMPF--LNPONGIMVNG 92
DB 11 YSFSANAPVFEYVPGGVVFTLDALGGSYD-----KIDFSKYNPATGPFVNG 60
QY 93 AEKGOVLAVYIESM-LPRGVDYXGICAMIPHFGLTGTDITAMINDPLPEKVMKILDE 151
DB 6: VKGCVIF-KGKIPAMPKMGVIGVAYD-EEVPTGPGKHGGMNDINLRKGTTFYFVF 106
QY 152 KYWYKRRH---TLPYKPHIGTILSVSPEDSINSLTPDNHGNMVDYDIPGSGITYPLVRA 208
DB 107 K--WAVLFCGVRPLPFWGVIGVAPQEGEYPTGTAHRHGNMDIKETENTVTHLPVQ 164
QY 209 PGHFLIGDAHACOGDGEICGTAVEFAFASITIKVDLIKNNQSWPRMENAINSIGRA 268
DB 165 EGALLAIGDGHATMGDCVCSACEVPAKVVFIDVSK-EIKWPVETNDAYYIIVSLP 223
QY 269 RPLDTRATRYARDLIYWLVEDEGFEQWDMYMLISQCGKVRGLNMVDPKIVTGAMLNK 324
DB 224 DDEALKVENTRETVMVLTORRKTIPP'DAYMIASLSVDGISLVNPAKTAARIPK 279

RESULT 11
Q8YVY6 PRELIMINARY; PRI: 328 AA.
AC Q8YVY6:
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 02-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein: All0706.
GN ALL0706.
OS Arabidopsis sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCB_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Katanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki M., Shimizu S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
KT cyanobacterium Anabaena sp. strain PCC 7120.";
KJ DNA Res. 8:205-213(2001);
DR EMBL: AF003583; BAB72663.1;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA.
DR Hypothetical protein: Complete proteome.
SQ SEQUENCE 328 AA: 35627 MW: DE5FBA0E4AF4165D CRC64;

Query Match 15.4%; Score 269; DB 16; Length 328;
Best Local Similarity 27.6%; Pred. No. 3.8e-15;
Matches 91; Conservative 59; Mismatches 140; Indels 40; Gaps 9;

QY 21 VTHLTEMOKEPHYTIGPVS---TPVLIEPCDRIIVDIRCAFEGAINSEQDIPSQLK 77
DB 1 MTHILKATKSTVE--IGGFSHJLEFALKVDSST--DVETTYG-----YVDKAPPELT 54
QY 78 MPFIN-----PQ-----NGPTVMGAFKGVLAIVIESMLPRGVDYXGICAMIP 121
DB 55 PEFIDICQLPPEKRIAGPHLLTGPVYKDAEPGDVLEVKLEAIAPS--LPVGNNAIRS 112
QY 122 HFGGITDITAMINDPLPEKVMKIL-----DSEKVVWYKRRHLPYKPHIGTILSVSPED 177
DB 113 GWG-----ALPNOFTPALRFILNLANNTAEPPNSGKIKPLTPPFGLGVATPEN 164
QY 176 SINSLTPDNHGNMVDYDIPGSGITYPLVRAPGRFLIGDAHACOGDGEICGTAVEFAFI 237
DB 165 ARSVPVPGYGGNIDNRELQAGSRIFLIPVPGGLFSLGDSGSAQGDGEVNTAIEISMN 224

```

```
QY 238 TTIKVLKKNQLSWPRMENENIMSGSARPLEDATRATAYRDLIYWVEDFGFEQDAY 297
DB 225 GRIQTLTKKJHFTAPTAETPTDIITIGFAPTDDAALEQALKNMIDFLRFVNLSPEDAY 284
QY 298 MLLSCGKVRIGNVMD-PKYTVGAMLNKL 326
DB 285 VLCSLAVNFRITOVVNSPKNGVHGLLPKL 314

RESULT 12
Q927N6
ID Q927N6 PRELIMINARY: PRT: 299 AA.
AC Q927N6;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DI 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein lin2752.
GN LIN2752.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Franquet L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Echoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurgey C.,
RA Etienne K.-D., Fstai H., Garcia-Jel Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kust F., Kuzapkat G.,
RA Madueno E., Mailounan A., Marz Vicente J., Ng E., Nodjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Nuez J.-C., Porcelli K.,
RA Remmel B., Rose M., Schluerer T., Simoes N., Tietz P.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:819-852(2001)
DR EMBL: AL596173; CAC97978.1;
DR ListList: LIN02752;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 299 AA: 3383 MW: 879878A0B90A70D9 CRC64;
Query Match 15.0%; Score 261.5; DB 16; Length 299;
Best Local Similarity 29.0%; Pred. No. 1.5e-14;
Matches 79; Conservative 44; Mismatches 112; Indels 37; Gaps 7;
QY 43 PVLTIPEGDRIIVDTDAFEAGINSEQDIPSQLKMPFLNPGCPIMVNGAEKGDVAVY 102
DB 20 PVLTVKDGSVVTKIKDHFNGQINKEQLHYGEIDWKQF-SPTGPICTEEAREFGDILAVT 78
QY 103 TESMLPRGVDPYGCAMIPHEFGTGTDTATMLNDP-----LPEK-VRMKILDSEKV 153
DB 79 IEKJT-----LTSKDV-VLLNGPTIGVTDLLPNCIRRYKIKONKI 119
QY 154 YNSKRTILPYKPHGTLSVSEIDNSL-----TPDNHGNNMVDPCPSITFPLVRAP 209
DB 120 IYSDEIHVQLOKTIQGL-----KTESIHQTFPHPIKGGTGLSPDITEGATIFPVEKF 175
QY 210 GGRLFIGDAHACQGDGEI--CGTAVEFASITTTKVDLIKNNQLSWPRMENENIMSGSAR 269
DB 176 GAMLHVAIDRATIGTKTITSAEVPAEVLRLQGLKWK-APTPII-HKNNLICLSFLT 235
QY 270 LEDATRIAYRDLIYWVEDFGFEQDAYMLLS 301
DB 236 TEKATKKALONMLNVMESDKITLEDAT-FLLS 267

RESULT 13
Q9C453
ID Q9C453 PRELIMINARY: PRT: 411 AA.
AC Q9C453;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DI 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Formamidase (EC 3.5.1.49).
GN FMDS
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MECLINE=21066161; PubMed=11139495;
RA Fraser J.A., Davis M.A., Hynes M.J.;
RT "The Formamidase Gene of Aspergillus nidulans: Regulation by Nitrogen
RT Metabolite Repression and Transcriptional Interference by an
RT Overlapping Upstream Gene.";
KL Genotics 157:119-131(2001).
DR EMBL: AF274009; AAG60585.1;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA; 1.
KW Hydrolase.
SQ SEQUENCE 411 AA: 44907 MW: D36B46A5243AF242 CRC64;
Query Match 13.7%; Score 239.5; DB 3; Length 411;
Best Local Similarity 25.5%; Pred. No. 1.9e-12;
Matches 95; Conservative 45; Mismatches 123; Indels 109; Gaps 17;
QY 30 OKFPHYTIGTYST-VLTIPEGDRIIVDTDAFEAGI--NSEQDIPS-QLLKMPFLNPNQ 85
DB 21 QKGLNRRHPPDIPSYGTIANNEVVKISCLDWTGGQIKNDSDADDIKNVDITQIHYL--S 77
QY 85 GPIVNGAEKGDVAVYTESMLPRGVDPYGCAMIPHEFGTGTDTATMLNDPPEKVRM 145
DB 78 GPFDIETAEFGDVLVLIQCVQPFEDQPFQVGFARENG-----GGFLEIYPPAKA 131
QY 146 TKLDSKVKYWSKRITLPKPHI-----GTLSVSP-----EIDNSITLP 184
DB 132 I-WDEGIFGSSRH----PHVREAGLIHPGLGCAPSAEVLAFNRRREGELIAANTLGR 186
QY 185 D-----NUGGNMVDPCPSITFPLVRAP 209
DB 187 EVAKPPSPSNVIAGSAGFLAAGTIGRGARTIPGRPHGNCNDIKNISGSKVILFVHPV 246
QY 210 GGRLFIGDAHACQGDGEI--CGTAVEFASITTTKVDLIKNNQLSWPRMENENIMSGSAR 257
DB 247 GAKFSVGDHFHFSQGDGEISFCG-AIEMAGVITLKTVKD---GMAKMAMKSPIFHPGV 302
QY 268 RP-----LEDATRIAYRDLIYWLVE---DFGFEQDAYMLLSQ 302
DB 303 EPQFGPGRYLTPEGFSVDEKQKQHYLDAT-VAYKQTCRLVIEYLRRYGNDYQIYLLS- 360
QY 303 CKVR--LGNKVD 313
DB 361 CAPVOGHIAGLVD 373

RESULT 14
Q9URY7
ID Q9URY7 PRELIMINARY: PRT: 410 AA.
AC Q9URY7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DI 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Formamidase-like protein.
GN SpAc869.04.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
```

